ARNOLD WHITE & DURKEE
PROFESSIONAL CORPORATION
ATTORNEYS AT LAW

Austin
Chicago
Houston
Menlo Park
Minneapolis
Washington

One Franklin Square
1301 K Street, N.W.
Suite 350 East Tower
Washington, DC 20005

Telephone 202.289.0229
Facsimile 202.289.7535

July 16, 1999

FILE: STFD:009--1

BOX PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

RE: *U.S. Patent Continuation Application Entitled: Method for Determining the Relative Abundances of Polynucleotide Sequences (Continuing Application of Serial No. 08/688,488 Filed July 30, 1996)*

Sir:

Transmitted herewith for filing are:

(1) Request for Filing Continuation Application Under 37 C.F.R. § 153(c).
(2) Information Disclosure Statement
(3) Copy of the Recorded Assignment for the Original Application
(4) Petition Under 37 C.F.R. § 1.182 to Change Order of Names of Inventors;
(5) Power of Attorney; and
(6) Preliminary Amendment.

The Assistant Commissioner is authorized to deduct said fees from or to Arnold White & Durkee Deposit Account No. 01-2508/STFD:009--1/RHO.

Please date stamp and return the enclosed postcard to evidence receipt of these materials.

Please forward any reply to this communication directly to our Houston office for docketing purposes. The mailing address is 750 Bering Drive, Houston, Texas, 77057, and the Houston fax number is 713.787.1440.

Respectfully submitted,

Glenn W. Rhodes w/permission DRA

Glenn W. Rhodes
Reg. No. 31,790

The PTO did not receive the following listed item(s) check 1,764

GWR:eb
Encl:

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Patrick O. Brown and Tidhar Dari Shalon	§	**Group Art Unit:** 1634
Serial No.:	Not yet assigned	§	**Examiner:** Ardin Marschel
Filed:	Herewith	§	**Attorney Docket No.:** STFD:009--1
For:	Method For Determining the Relative Abundances of Polynucleotide Sequences	§	(Continuing application of Serial No. 08/688,488, filed July 30, 1996)

EXPRESS MAIL MAILING LABEL

NUMBER: EM540830375US
DATE OF DEPOSIT: November 24, 1998

I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.

Emily A. Evans
Signature

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Please amend the above-identified application as follows:

AMENDMENT

In the Title

Please change the title to -- Substrates Comprising Polynucleotide Microarrays --.

In the Specification

At page 1, immediately beneath the title of the invention, please insert the following:

-- The United States government may have certain rights in the present invention pursuant to Grant No. HG00450 by the National Institutes of Health. --.

At page 1, line 4, after "invention is a" insert -- continuation of U.S. patent application Serial No. 08/688,488 for "Method For Analyzing Gene Expression Patterns", filed July 30, 1996, which is --.

At page 1, line 10, after "1995," insert -- now U.S. patent No. 5,807,522, --.

At page 1, line 14, after "1994" insert --, now abandoned --.

At page 1, line 14, delete "three" and substitute therefor -- four --.

At page 3, line 27, delete "twofold" and substitute therefor -- two-fold --.

At page 4, line 12, delete "from"; and

line 35, delete "filters" and substitute therefor -- filters, --.

At page 5, line 12, delete "abundances" and substitute therefor -- abundances, --; and

line 27, delete "locations" and substitute therefor -- locations on --.

At page 10, line 5, after "100/cm^2," insert -- for example about 400/cm^2, --.

At page 10, line 8, after "10-500 μm," insert -- for example about 250 μm, --.

At page 11, line 20, delete "patents" and substitute therefor -- patent --; and

line 21, delete "applications" and substitute therefor -- application --.

At page 13, line 4, delete "OH" and substitute therefor -- hydroxyl --.

At page 14, line 5, delete "number" and substitute therefor -- number, --.

At page 16, line 23, delete "mixture" and substitute therefor -- mixtures --.

At page 17, line 29, delete "of" and substitute therefor -- of the --.

At page 18, line 20, delete "in changes of" and substitute therefor -- of changes in --.

At page 19, line 25, delete "*cerivisiae*" and substitute therefor -- *cerevisiae*--; and

line 29, delete "labeled two" and substitute therefor -- two labeled --.

At page 20, line 32, delete "small" and substitute therefor -- a small --.

At page 21, line 34, delete "wildtype" and substitute therefor -- wild type --; and

line 35, delete "*arabidopsis*" and substitute therefor -- *Arabidopsis* --.

At page 22, line 4, delete "*arabidopsis*" and substitute therefor -- *Arabidopsis* --;

line 6, delete "wildtype" and substitute therefor -- wild type --;

line 15, delete "*arabidopsis*" and substitute therefor -- *Arabidopsis* --; and

line 33, delete "or" and substitute therefor -- of --.

At page 23, line 6, delete "*arabidopsis*" and substitute therefor -- *Arabidopsis* --.

At page 24, line 22, delete "of distal end" and substitute therefor -- of the distal end of --.

At page 26, line 13, delete "were used as probes" and substitute therefor -- was used as a probe --; and

line 16, delete "analysis," and substitute therefor --analysis--.

At page 28, line 16, delete "activator" and substitute therefor -- activator of --;

line 22, delete "were" and substitute therefor -- was --; and

line 32, delete "Arabidopsis" and substitute therefor -- *Arabidopsis* --.

At page 29, line 30, delete "suggest" and substitute therefor -- suggests --.

At page 30, line 10, delete "tissues" and substitute therefor -- tissue --.

At page 31, line 15, delete "two color" and substitute therefor -- two-color --;

line 18, delete "samples" and substitute therefor -- samples will --; and

line 20, delete "genes responses to heatshock" and substitute therefor -- genes' responses to heat shock --.

At page 32, line 26, delete "H2O" and substitute therefor -- H_2O --;

line 26, delete "min" and substitute therefor -- minute --;

line 27, delete "NaBH4" and substitute therefor -- $NaBH_4$ --;

line 29, delete "H2O" and substitute therefor -- H_2O --;

line 29, delete "min" and substitute therefor -- minutes --;

line 30, delete "min" and substitute therefor -- minute --; and

line 31, delete "H2O" and substitute therefor -- H_2O --.

At page 33, line 15, delete "min" and substitute therefor -- minutes --;

line 17, delete "min" and substitute therefor -- minutes --;

line 24, delete "speedvac" and substitute therefor -- SPEEDVAC --;

line 25, delete "purchase" and substitute therefor -- purchased --;

line 32, delete "sulfate;" and substitute therefor -- sulfate --; and

line 33, delete "SDS)." and substitute therefor -- (SDS)). --.

At page 34, line 5, delete "hr" and substitute therefor -- hours --;

line 6, delete "min" and substitute therefor -- minutes --; and

line 8, delete "min" and substitute therefor -- minutes --.

At page 35, line 14, delete "acids derived" and substitute therefor -- acid-derived --.

At page 37, line 34, delete "0.1%SDS" and substitute therefor -- 0.1% SDS --; and

line 34, delete "cover slipped" and substitute therefor -- cover-slipped --.

At page 38, line 31, delete "above" and substitute therefor -- above, --.

At page 39, line 16, delete "polyA+" and substitute therefor -- poly A+ --.

At page 41, line 4, delete "micro array" and substitute therefor -- microarray --;

line 6, delete "acetylcholine" and substitute therefor -- acetylcholine receptor --;

line 22, delete "Inserts" and substitute therefor -- Inserts of --; and

line 25, delete "E. coli" and substitute therefor -- *E. coli* --.

At page 42, line 2, delete "Arabidopsis" and substitute therefor -- *Arabidopsis* --;

line 10, delete "CO2" and substitute therefor -- CO_2 --;

line 21, delete "Arabidopsis" and substitute therefor -- *Arabidopsis* --;

line 26, delete "1,1,000" and substitute therefor -- 1:1,000 --; and

line 34, delete "were" and substitute therefor -- was --.

At page 43, line 6, delete "Arabidopsis" and substitute there for -- *Arabidopsis* --;

line 24, delete "polyA+" and substitute therefor -- poly A+ --; and

line 31, delete P32-dATP" and substitute therefor -- P^{32}-dATP --.

At page 44, line 10, delete "used" and substitute therefor -- used as the --.

In the Claims

Please cancel claims 1-6 without prejudice.

Please add claims 7-37 as follows:

--7. A substrate with a surface comprising a microarray of DNA sequences, wherein (i) the microarray has a density of about 400 or more discrete regions of DNA sequences per cm^2 of substrate surface, (ii) the DNA sequences are isolated polynucleotides, (iii) the microarray comprises 400 or more regions, and (iv) each region in the microarray is free of cross-contamination with DNA sequences applied to the other regions in the microarray.--

-- 8. The substrate of claim 7, wherein the density of discrete regions of DNA sequences is at least about 10,000/ cm^2. --

-- 9. The substrate of claim 7, wherein the density of discrete regions of DNA sequences is at least about 2500/ cm^2. --

-- 10. The substrate of any one of claims 7, 8 or 9, wherein the substrate is glass. --

-- 11. The substrate of any one of claims 7, 8 or 9, wherein the substrate is non-porous. --

-- 12. The substrate of any one of claims 7, 8 or 9, wherein the surface of the substrate is hydrophobic. --

-- 13. The substrate of any one of claims 7, 8 or 9, wherein the surface of the substrate comprises one or more chemical moieties selected from the group consisting of silyl, hydroxyl, carboxyl, amine, aldehyde, and sulfhydryl. --

-- 14. The substrate of any one of claims 7, 8 or 9, wherein the DNA sequences are

covalently bound to the surface of the substrate. --

-- 15. The substrate of any one of claims 7, 8 or 9, wherein the DNA sequences are non-covalently bound to the surface of the substrate. --

-- 16. The substrate of claim 15, wherein the DNA sequences are non-covalently bound to a polycationic polymer on the surface of the substrate. --

--17. The substrate of any one of claims 7, 8 or 9, wherein the DNA sequences are selected from the group of polynucleotides consisting of mRNA-derived sequences, genomic DNA sequences and fragments thereof.--

-- 18. The substrate of any one of claims 7, 8 or 9, wherein the microarray comprises 2,500 or more regions. --

-- 19. The substrate of any one of claims 7, 8, or 9, wherein the microarray comprises 10,000 or more regions. --

-- 20. The substrate of any one of claims 7, 8, or 9, wherein the DNA sequences are bound directly to the surface of the substrate. --

-- 21. A substrate with a surface comprising a microarray of DNA sequences, wherein the DNA sequences are polynucleotides, produced by a method comprising the steps of

(a) depositing a selected volume between about 0.002 nl and about 2 nl of a solution comprising a selected, isolated polynucleotide at a discrete region on the surface of the substrate, and

(b) repeating step (a) at other locations on the surface of the substrate until a microarray of 400 or more regions is formed, wherein the regions are at a density between about 62,500 regions/cm^2 and about 625 regions/cm^2. --

-- 22. The substrate of claim 21, wherein the density of discrete regions in the microarray is about 10,000 regions/cm^2 or more. --

-- 23. The substrate of claim 21, wherein the density of discrete regions in the microarray is about 2,500 regions/cm^2 or more. --

-- 24. The substrate of claim 21, wherein the substrate is glass. --

-- 25. The substrate of claim 21, wherein the substrate is non-porous. --

-- 26. The substrate of claim 21, wherein the surface of the substrate is hydrophobic. --

-- 27. The substrate of claim 21, wherein the surface of the substrate comprises one or more chemical moieties selected from the group consisting of silyl, hydroxyl, carboxyl, amine, aldehyde, and sulfhydryl. --

-- 28. The substrate of claim 21, wherein the DNA sequences are bound directly to the surface of the substrate. --

-- 29. The substrate of claim 21, wherein the DNA sequences are covalently bound to the surface of the substrate. --

-- 30. The substrate of claim 21, wherein the DNA sequences are non-covalently bound to the surface of the substrate. --

-- 31. The substrate of claim 21, wherein the DNA sequences are selected from the group of polynucleotides consisting of mRNA-derived sequences, genomic DNA sequences, and fragments thereof. --

-- 32. The substrate of any one of claims 21, 22 or 23, wherein the microarray comprises 2,500 or more regions. --

-- 33. The substrate of any one of claims 21, 22, or 23, wherein the microarray comprises 10,000 or more regions. --

-- 34. A substrate with a surface comprising a microarray of DNA sequences and suitable for analysis of a polynucleotide mixture, wherein (i) the microarray has a density of about 400 or more discrete regions of DNA sequences per cm^2 of substrate surface; (ii) each of said regions contains, as an isolated polynucleotide, a characteristic DNA sequence; (iii) the microarray comprises at least 400 regions free of cross-contamination by DNA sequences characteristic of others of said 400 regions. such that the DNA sequences in said regions are selective in hybridizing with corresponding members of said mixture. --

-- 35. A substrate according to claim 34, wherein the freedom from cross-contamination in said 400 regions is sufficient to permit detection of a two-fold change in the relative abundance of polynucleotides in mixtures subjected to analysis. --

-- 36. A substrate with a surface comprising a microarray of DNA sequences, wherein (i) the microarray has a density of 400 or more discrete regions of DNA sequences per cm^2 of

substrate surface, (ii) the DNA sequences are purified polynucleotides, and (iii) the DNA sequences are non-covalently bound to a polycationic polymer on the surface of the substrate --.

-- 37. The substrate according to claim 36, wherein the DNA sequences are cDNAs --.

REMARKS

The title has been amended to make it more descriptive of the claimed invention. The specification has been amended to reflect that the United States government may have certain rights in the present invention. The specification has been amended to reflect that this application is a continuation of application Serial No. 08/688,488, and to update the status of the great grandparent and great-great-grandparent applications. The specification has been amended to correct various typographical and syntax errors. It is apparent that no new matter was added by any of these amendments and they require no further comment.

Applicants have amended the specification to describe the present invention with greater particularity. Specifically, Applicants have amended the specification explicitly to recite that the present microarray invention includes microarrays having a density of at least about 400 per cm^2, corresponding to typical diameters of about 250 μm for regions in the microarray. Support for these amendments to the specification is found, among other places, in priority application Serial Nos. 08/261,388 (great-great-grandparent) and 08/477,809 (great-grandparent), which were incorporated by reference in the present application as filed at page 1, lines 4-15. For example, the great-great-grandparent application recites at page 8, lines 7-11, that each region in the microarray may be "about 25-**250** μm, and are separated from other regions in the array by about the same

With respect to the limitation in claim 35 regarding detection of a two-fold change in polynucleotide abundance, Applicants refer the Examiner to page 8, lines 13-19, and page 26, lines 20-23 of the application as filed.

New claim 36 is directed to a particular embodiment of Applicants' invention wherein (1) the microarray has a density of 400 or more regions per cm^2 of substrate surface, (2) the DNA sequences at each region of the microarray are purified polynucleotides and (3) the DNA sequences are non-covalently bound to a polycationic polymer on the surface of the substrate. The support for the density limitation is as described by Applicants for claim 7. The support for the non-covalent binding to a polycationic polymer is the same as for claims 15 and 16. *See, e.g.*, page 13, lines 15-23, of the specification. The support for the DNA sequences being purified polynucleotides is found at, among other places, page 9, line 21 to page 10, line 9 and page 14, line 17 to page 15, line 14. By the term "purified DNA sequences", Applicants mean that the DNA sequences are not present in a cell. New dependent claim 37 is supported, for example, by the disclosure at page 15, lines 3-7.

Applicants note that an Information Disclosure Statement is being filed concurrently with this Preliminary Amendment. In addition, Applicants wish to apprise the Examiner of three related, pending applications, in addition to priority application Serial Nos. 08/261,388, 08/477,809 (now U.S.P. 5,807,522), 08/514,875, and 08/688,488. These are: (1) Serial No. 09/001,027, a continuation application of great-grandparent Serial No. 08/477,809, (2) an application filed November 19, 1998, Serial No. not yet assigned, which is a continuation of

application Serial No. 09/001,027, and (3) an application filed November 10, 1998, Serial No. not yet assigned. which is a continuation of grandparent Serial No. 08/514,875.

* * *

In view of the foregoing, entry of the amendments and allowance of the claims is respectfully requested. The Examiner is invited to contact the undersigned attorney at (650) 614-4654 with any questions, comments or suggestions relating to the above-identified patent application.

Respectfully submitted,

Emily A. Evans
Reg. No. 33,493

Attorney for Applicants

ARNOLD, WHITE & DURKEE
P.O. Box 4433
Houston, TX 77210
(650) 614-4654

Date: November 24, 1998

Attorney Docket No.: 8600-0138.32 Patent

Transmittal of Utility Patent Application for Filing

Certification Under 37 C.F.R. §1.10 (if applicable)

EM 383 301 715 US
"Express Mail" Label Number

July 30, 1996
Date of Deposit

I hereby certify that this Transmittal letter, enclosed application, and any other documents referred to as enclosed herein are being deposited in an envelope with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR §1.10 on the date indicated above and addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231

JANETTE CABAUTAN
(Print Name of Person Mailing Application)

(Signature of Person Mailing Application)

METHOD FOR DETERMINING THE RELATIVE ABUNDANCES OF POLYNUCLEOTIDE SEQUENCES

The present invention is a continuation-in-part of U.S. patent application Serial No. 08/514,875 for "Method and Gene-Array Device for Analyzing Gene Expression Patterns", filed August 14, 1995, which is a continuation-in-part of U.S. patent application Serial No. 08/477,809 for "Method and Apparatus for Fabricating Microarrays of Biological Samples", filed June 7, 1995, which is a continuation-in-part of U.S. patent application Serial No. 08/261,388 for "Method and Apparatus for Fabricating Microarrays of Biological Samples", filed June 17, 1994. These three applications are incorporated herein by reference.

Field of the Invention

This invention relates to a method for determining the relative abundances of a plurality of polynucleotide sequences.

References

Adams, M.D., *et al.*, *Science* 252:1651 (1991).

Adams, M.D., *et al.*, *Nature* 355:632 (1992).

Ausubel, F.M., *et al.*, Eds., in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene & Wiley Interscience, New York, NY (1994).

Baeuerle, P.A., *et al.*, *Annu. Rev. Immunol.* 12:141 (1994).

Becker, J., and Craig, E.A., *Euro. J. Biochem.* 219:11 (1994).

Bohlander, *et al.*, *Genomics* 13:1322 (1992).

Chan, A.C., *et al.*, *Annu. Rev. Immunol.* 12:555 (1994).

Cohen, G.B., *et al.*, *Cell* 80:237 (1995).

Collins, F.S., *Proc. Natl. Acad. Sci. USA* 92:10821 (1995).

Crabtree, G.R., *et al.*, *Annu. Rev. Biochem.* 63:1045 (1994).

Craig, E.A., *et al.*, *Cell* 78:365 (1994).

Cyr, D.M., *Trends Biochem. Sci.* 19:176 (1994).

Jakob, U., *et al.*, *Trends Biochem. Sci.* 19:205 (1994).

Jindal, S., *Trends Biotechnol.* 14:17 (1996).

Lehrach, H., *et al.*, "Hybridization Fingerprinting in Genome Mapping and Sequencing," in GENOME ANALYSIS VOLUME 1: GENETIC AND PHYSICAL MAPPING (Davies, K.E., and Tilgham, S.M., Eds.) Cold Spring Harbor Laboratory Press, (Cold Spring Harbor, NY) pp. 39-80 (1990).

Liou H.-C., *et al.*, *Curr. Op. Cell Biol.* 5:477 (1993).

Mullis, K.B., *et al.*, U.S. Patent No. 4,683,195, issued 28 July 1987.

Nelson, S.F., *et al.*, *Nature Genetics* 4:11-17 (1993).

Newton, A.C., *J. Biol. Chem.* 270:28495 (1995).

Nishizuka, Y., *FASEB J.* 9:484 (1995).

Riles, *et al.*, *Genetics* 134:81 (1993).

Rohan, P.J., *et al.*, *Science* 259:1763 (1993).

Sambrook, J., *et al.*, in MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition, Cold Spring Harbor Laboratory (Cold Spring Harbor, NY) (1989).

Schena, M., *et al.*, *Science* 270:467-470 (1995).
Shalon, D., thesis, Stanford University (1995).
Thanos, D., *et al.*, *Cell* 80:529 (1995).
Wilkinson, K.D., *Annu. Rev. Nutr.* 15:161 (1995).

Background of the Invention

With a large number of human genes now identified through the Human Genome Project, there is great interest in finding out how these genes act in concert to regulate the whole organism (Watson, Collins, 1995; Adams, *et al.*, 1991, 1992; Cohen, *et al.*, 1995; Chan, *et al.*, 1994; Crabtree, *et al.*, 1994).

The current focus of this research is in monitoring the genes' activities, *i.e.*, levels of expression, as a function of cell type, cell condition, disease state, or drug therapy response. The general requirements of this type are severalfold:

First, the method must be able to handle large numbers of genes at once. Ideally, all or nearly all expressed genes from a given cell type may be represented.

Secondly, the method should be responsive to relatively small, as well as to large, changes in the levels of expression. For example, in monitoring the response of genes in a cell exposed to a given drug, it may be important to detect slight changes in levels of expression of genes, *i.e.*, a twofold increase or decrease in expression level, in order to identify drug-responsive genes. As another example, in studying the response of genes in a given disease state, *e.g.*, tumor state, it may be important, in understanding the relationship between gene expression and disease state, to classify genes according to low, moderate, and high shifts in gene expression. More generally, the greater resolution that can be achieved, in terms of relative levels of expres-

sion, the more information that can be obtained from the studies.

Finally, the method should be amenable to small amounts of polynucleotide sample material, to obviate the need for amplification of sample material, with the attendant possibility of differential amplification.

A variety of methods for studying changes in gene expression have been proposed heretofore. One approach is based on differential hybridization between nucleic acid fractions from control and test sources. After hybridization of cDNA's from the two sources, those species from "equally expressed" can be removed as DNA hybrids, leaving overexpressed or underexpressed cDNA's in single-stranded form. The single-stranded species can then be further characterized, *e.g.*, by electrophoretic fractionation.

This approach has been useful as a starting point for isolation or characterization of polynucleotides of interest in a polynucleotide mixture, but is not suited to following small changes in expression levels, or tracking changes in a relatively large number of genes.

Another approach for comparing the relative abundances of polynucleotides in mixtures involves hybridizing individual labeled probe mixtures with different replicate filters containing immobilized gene sequences, *e.g.*, colonies of cloned DNA. Colonies on the replicate filters which hybridize differentially to the two probe mixtures then represent gene sequences which are present in greater or lesser abundance in the two probe mixtures.

Because this method relies on comparing the amounts of label at corresponding positions on two different substrates (or filters), the resolution of the method is limited by (i) variations in the amount of immobilized DNA at corresponding array positions on the two different filters (ii) variations in the extent of hybridization

that occurs at corresponding array position, *e.g.*, due to differences in probe accessibility to the immobilized DNA, and (iii) variations in measured reporter levels at corresponding array positions. To improve the resolution appreciably, one would have to average out these variations by conducting many hybridization measurements in parallel.

It would thus be useful to provide, for measuring the relative abundances of polynucleotides in complex mixtures, an improved method that avoids or overcomes the limitations above. In particular, the method should be adaptable to measuring the relative abundances copy numbers or expression levels in a very large number (*e.g.*, 50,000 to 100,000) of genes, at a high resolution, *e.g.*, two-fold change in relative abundance, and at high sensitivity for rare genes. At the same time, the method should be simple to carry out.

Summary of the Invention

The present invention is directed to a method of determining the relative amounts of individual polynucleotides in a complex mixture of different-sequence polynucleotides. The method includes labeling the polynucleotides with a fluorescent reporter, and contacting the labeled polynucleotides, under hybridization conditions, with an array of different DNA sequences disposed at discrete locations a non-porous surface, at a density of at least about 100 sequences/cm^2. The different DNA sequences in the array are each present in multiple copies, and are effective to hybridize to individual polynucleotides in the mixture. The level of fluorescence at each position in the microarray is then determined.

The labeled polynucleotides are preferably contacted with the microarray by covering the array surface with a

solution of the mixture of labeled polynucleotides, to a solution depth of less than 500 microns.

In various preferred embodiments, the density of array elements corresponding to different-sequence DNA locations in the array is at least 1,000/cm^2, the DNA sequences in the array are at least about 50 bases in length, and the labeled polynucleotides represent at least 1 million unique base sequences.

The method may be used in determining the relative amounts of each polynucleotide from first and second different sources. Here (i) the polynucleotides from the first and second sources are labeled with independently detectable first and second fluorescent reporters, respectively, (ii) the contacting of labeled polynucleotides from first and second sources is carried out simultaneously under competitive hybridization conditions, and (iii) the determining step includes measuring the levels of the two reporters at each position in the array.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying figures.

Brief Description of the Drawings

Fig. 1 shows a portion of a two-dimensional microarray of different DNA sequences used in practicing the method of the invention;

Fig. 2 shows a fluorescent image of an actual 20 × 20 array of 400 fluorescently-labeled DNA samples immobilized on a poly-l-lysine coated slide, where the total area covered by the 400 element array is 16 square millimeters;

Fig. 3A is a fluorescent image of a 1.8 cm × 1.8 cm microarray containing lambda clones with yeast inserts,

the fluorescent signal arising from the hybridization to the array with approximately half the yeast genome labeled with a green fluorophore and the other half with a red fluorophore;

Fig. 3B shows the translation of the hybridization image of Fig. 3A into a karyotype of the yeast genome, based on the hybridization pattern of the Fig. 3A microarray which contains yeast DNA sequences that have been previously physically mapped in the yeast genome;

Figs. 4A and 4B show scans of hybridization signals from an array of genes probed with fluorescently-labeled *Arabidopsis* wild-type (4A) or transgenic HAT4 (4B) cDNA at low photomultiplier tube settings;

Figs. 5A and 5B show scans of hybridization signals from an array of genes probed with fluorescently-labeled *Arabidopsis* wild-type root (5A) or wild-type leaf (5B) cDNA at intermediate photomultiplier tube settings;

Fig. 6 shows a combined two-color scan of a microarray containing 1,046 cDNA's from peripheral blood lymphocytes (PBL's) after hybridization with a mixture of cDNA's from bone marrow labeled with Cy5-dCTP and cDNA's from Jurkat cells labeled with fluorescein-dCTP, where red spots indicate greater gene expression levels in bone marrow, green spots, greater gene expression levels in Jurkat cells, and yellow spots, comparable gene expression levels in both cells;

Fig. 7 shows a combined two-color scan of a microarray containing 1,046 cDNA's from peripheral blood lymphocytes (PBL's) after hybridization with a mixture of cDNA's from heat-shocked Jurkat cells labeled with Cy5-dCTP and cDNA's from control Jurkat cells labeled with fluorescein-dCTP, where red spots indicate greater gene expression levels in heat-shocked cells, green spots, greater gene expression levels in unshocked cells, and

yellow spots, comparable expression gene expression levels in both cells;

Fig. 8 shows a combined two-color scan of a micro-array containing 1,046 cDNA's from peripheral blood lymphocytes (PBL's) after hybridization with a mixture of cDNA's from phorbol-ester-treated Jurkat cells labeled with Cy5-dCTP and cDNA's from control Jurkat cells labeled with fluorescein-dCTP, where red spots indicate greater gene expression levels in phorbol-ester treated cells, green spots, greater gene expression levels in untreated cells, and yellow spots, comparable expression gene expression levels in both cells;

Figs. 9A and 9B are schematic displays of activated and repressed genes in Jurkat cells in response to heat shock (9A) and phorbol ester (9B), where the colors indicated on the display correspond to array elements that display greater than 2-fold elevation (red), less than a 2-fold change (black), or less than 2-fold repression (green);

Figs. 10A-10L are Northern RNA "dot" blots of samples of RNAs from control Jurkat cells (-HS) or heat-shocked Jurkat cells (+HS), after spotting onto nylon membranes, and blotting with the designated cDNA probes from Table 1;

Figs. 11A-11E are Northern RNA "dot" blots of samples of RNAs from control Jurkat cells (-PMA) or phorbol-ester-treated Jurkat cells (+PMA), after spotting onto nylon membranes, and blotting with the designated cDNA probes from Table 1; and

Figs. 12A-12C show transcript profiles of heat shock- and phorbol ester-regulated genes, where gene expression levels per 10^5 mRNAs (x-axes) are shown for 15 genes (Table 1) in human bone marrow, brain, prostate, and heart, and the genes are grouped according to expression levels (A-C).

Detailed Description of the Invention

I. Definitions

Unless indicated otherwise, the terms defined below have the following meanings:

A "polynucleotide" refers to a DNA or RNA polymer at least about 50 bases in length, *i.e.*, containing at least about 50 nucleotide subunits.

"Different-sequence polynucleotides" refers to polynucleotides having different, unique base sequences. Such polynucleotides are distinguished by their ability to hybridize selectively to complementary-strand nucleic acid sequences under selected hybridization conditions.

A "complex mixture of different-sequence polynucleotides" refers to a mixture containing a plurality, and preferably at least 100, different-sequence polynucleotides. The complexity of the mixture is defined by the aggregate unique base sequence in the different-sequence polynucleotides. For example, a mixture of 1,000 polynucleotides, each containing 1,000 unique base sequences, will have a complexity of 1 million bases.

A "microarray of DNA sequences" refers to a spatial array of nucleic acid polymers, *e.g.*, polymers of at least 15-20 bases, preferably polynucleotides of 50 bases or more, having a different DNA sequence at the different microarray locations. The different sequences may be of known sequence, or partially or completely sequenced, as with polymers prepared by solid-phase synthesis, primer-amplified genomic DNA fragments or expressed-sequence tags (EST clones) (Adams, *et al.*, 1991, 1992), or they may be unsequenced, as with library cDNA's isolated from a library. The array may also contain regions with different graded concentrations or sequence lengths of same-sequence polynucleotides, and/or mixtures of two or more different sequences. The different DNA sequences may also include DNA analogs, such as analogs having

phosphonate or phosphoramidate backbones, capable of hybridizing, through Watson-Crick base pairing, with complementary-sequence DNA or RNA. The microarray has a density of distinct gene sequences of at least about 100/cm^2, and preferably at least about 1,000/cm^2. The regions in a microarray have typical dimensions, *e.g.*, diameters, in the range between about 10-500 μm, and are separated from other regions in the same array by about the same distance.

"Cells of a given cell type or types" refers to cells obtained from one or more particular tissues or organs, *e.g.*, hepatocytes, heart muscle cells, pancreatic cells, or non-differentiated embryonic tissue, or to a particular blood cell type or types, *e.g.*, peripheral blood lymphocytes.

Cells having a "selected physiological state or disease condition" or "test cells" refer to cells of a given cell type or types which, as examples, (i) are in a defined state of differentiation or activation, *e.g.*, by gene activation; (ii) are infected by a defined infectious agent, *e.g.*, HIV-infected T cells; (iii) are in a neoplastic state, *i.e.*, tumor cells; (iv) are in a chemical- or physical-response state, *i.e.*, after exposure to a pharmacological agent with respect to control cells of the same type or types; (v) are in a defined stage of cell cycle; (vi) have a particular chronological age; (vii) are undergoing response to an extracellular signal; (viii) have a genetic defect; (ix) are from a patient with a particular physiological disease state; or (x) are taken from particular anatomical locations or microenvironments.

Cells which are in a normal or control or an alternative reference state with respect to "test cells" are referred to herein as "control cells".

"Reporter-labeled copies of messenger nucleic acid" refers to reporter-labeled mRNA transcripts obtained from test or control cells or cDNAs produced from such transcripts. The reporter label is any detectable reporter, and typically a fluorescent reporter.

A "cDNA" refers to a cloned, amplified, synthesized, single-stranded, double-stranded, or first strand product DNA corresponding in sequence to all or a portion of an mRNA.

II. DNA Sequence Microarrays

This section describes microarrays of different DNA sequences disposed at discrete locations on a non-porous surface, at a density of at least about 100 sequences/cm^2. The DNA sequences in the array are each (i) present in multiple copies, and (ii) effective to hybridize to an individual polynucleotide in a mixture of polynucleotides, in accordance with the method of the invention.

Methods of forming microarrays of this type have been detailed in the above-cited, co-owned U.S. patents applications Serial Nos. 08/514,875, 08/477,809, and 08/261,388, and related PCT application WO/95/35505, published December 28, 1995, all of which are incorporated by reference herein. Details are also given in the Methods below.

Briefly, a capillary- or tweezer-like fluid dispenser in a robotic apparatus is employed to pick up a selected DNA solution from a solution source, and distribute the solution to each of a plurality of non-porous microarray substrates, at a selected microarray location on the substrate surface, and in a selected amount (solution volume). A droplet of solution is deposited on the substrate surface by tapping the dispenser on the substrate at the selected deposition region. The volume of material deposited can be controlled according to the

dimensions of the dispenser, the tapping force, the viscosity of solution, and the duration of application, to achieve a selected deposition volume, as described in the above co-owned patent applications. Typically deposition volumes are between about 2×10^{-3} to 2 nanoliters (nl), corresponding in droplet size (surface diameter) between about 20 to 200 μm. The DNA in the each microarray region is preferably present in a defined amount between about 0.1 femtomoles and 100 nanomoles.

This operation is repeated for each microarray position, until the array is completed, with a different DNA sequence at each position (recognizing that some microarray positions may be duplicates with the same or different amounts of a single sequence, and some microarray positions may contain combinations of two or more different DNA sequences).

As noted above in the definitions above, the regions of the microarray are about 20-500 μm in diameter, preferably about 50-200 μm, with spacing between adjacent regions on the microarray surface of about the same dimensions. The microarray density is at least 100/cm^2, and preferably at least 1,000/cm^2. Thus, for example, an array having 100 μm size DNA spots, each separated by 100 μm, will have a density of about 2,500 regions/cm^2.

The non-porous substrate surface on which the different DNA sequences are deposited may be any surface capable of supporting an aqueous layer on the surface without liquid absorption or flow into or through the substrate; that is, any substrate surface which is water-impermeable. Preferred substrates are glass slides, metalized surfaces, non-porous polymer substrates, such as polyethylene, polyurethane, or polystyrene sheet material, or substrates which are coated with a non-permeable film or layer.

In the case where the DNA sequences are to be covalently attached to the substrate surface, the substrate includes or is treated to include chemical groups, such as silylated glass, OH, carboxyl, amine, aldehyde, or sulfhydryl groups. After deposition of the DNA sequences on the substrate surface, and either before or after drying the solution at each array location, the DNA is fixed by covalent attachment to the surface. This may be done, for example, by drying the DNA spots on the array surface, and exposing the surface to a solution of a cross-linking agent, such as glutaraldehyde, borohydride, or any of a number of available bifunctional agents. One such linking method is detailed in the Methods section below.

Alternatively, the DNA sequences may be attached to the substrate surface non-covalently, and typically by electrostatic interaction between positively charged surface groups and the negatively charged DNA molecules. In one preferred embodiment, the substrate is a glass slide having formed on its surface, a coating of a polycationic polymer, preferably a cationic polypeptide, such as polylysine or polyarginine, as described in the Methods section below. In experiments conducted in support of the invention, as illustrated in Example 1 and 2, it has been discovered that the non-covalently bound polynucleotides remain bound to the coated slide surface when an aqueous DNA sample is applied to the substrate under conditions which allow hybridization of reporter-labeled polynucleotides in the sample to complementary-sequence (single-stranded) polynucleotides in the substrate array.

The DNA sequences forming the microarray are polynucleotides having lengths of about 50 bases or greater, although shorter DNA polymers, *e.g.*, in the range 15-50 bases may also be employed. These DNA sequences can be

formed by a variety of methods, including solid-phase synthesis, polymerase chain reaction (PCR) methods, gene cloning, gene cloning in combination with PCR, and solid-phase methods. The DNA sequences on the array have the same sequence as some, and preferably a large number of the polynucleotides in the complex mixture which will be reacted with the microarray. That is, the microarray sequences will hybridize with selected polynucleotides in the polynucleotide mixture. Moreover, the hybridization is selective, meaning that one array sequence will typically hybridize with one species only in the polynucleotide mixture. This requirement would exclude relatively short, *e.g.*, less than 8-10 base oligonucleotides as array DNA sequences and/or combinatorial oligonucleotides such as are used for sequencing by hybridization.

Preferred sources of DNA sequences in the microarrays are genomic sequences and mRNA-derived sequences, as illustrated in the examples below. DNA sequences from genomic sources are typically obtained from cloned genomic fragments corresponding predominantly to single-copy or low-copy number genes, *i*.e., excluding repeat-sequence genomic fragments, or by primer-amplification of transcribed regions of a genome.

The cloned fragments may be excised and/or PCR-amplified, then purified by conventional methods, such as described in the Methods below. Alternatively the fragments may be obtained directly from genomic sources, *e.g.*, from genomic fragments that have been treated to remove repeat sequences, and/or by PCR amplification methods. For example, using computer-aided sequence analysis, it is possible to construct PCR primers for expressed genomic sequences, to selectively generate these sequences. The purified genomic sequences applied

to the microarray surface may be partially or completely sequenced, or may have unknown sequences.

DNA sequences derived from mRNA's can include mRNA's themselves, but more commonly sequences produced by reverse transcription of mRNA's, or DNA sequences obtained from a library of cloned cDNA's, *e.g.*, by excision or primer amplification. The mRNA's are preferably obtained from cells having a selected physiological state or disease condition and from corresponding test cells, as defined above, according to known methods. The mRNA-derived sequences applied to the microarray surface may be completely sequenced, partially sequenced, *e.g.*, as for expressed sequence tags (EST's), or may have unknown sequences.

Fig. 1 shows a portion of a microarray 20 having a substrate 22 whose surface 24 contains an array of regions, such as at 26, containing different-sequence DNA's. A 20 × 20 array in a $(4mm)^2$ area, and thus having a density of about 2,500 regions/cm^2 is shown in Fig. 2. The array is formed as described in Example 1.

III. Microarray Method

The method of the invention is designed to measure the relative abundances of polynucleotides in a complex mixture of polynucleotides. As defined above, a complex mixture preferably contains at least about 50, and up to 1,000 or more polynucleotides with different sequences. The complexity of the mixture is defined by the number of unique polynucleotide sequences and is preferably at least 10^6, e.g., containing 1000 different-sequence polynucleotides having an average of at least 1,000 bases.

Genomic fragments of the type described in Section II above represent one source of polynucleotides, where the method is used to determine the relative abundance of

different genomic species, *e.g.*, from a selected chromosome or region of a chromosome, or chromosomes of a given cell type. Methods of obtaining genomic fragments suitable for the method are described above, or are known, *e.g.*, as described in U.S. Patent No. 5,376,526.

Polynucleotides derived from mRNA's as above represent another source of polynucleotides, where the method is used to determine the relative abundance of mRNA's from a given cell source, for purposes of measuring the relative levels of gene expression of a plurality of genes, *e.g.*, from the cell source. Methods for obtaining mRNA-derived polynucleotides, *e.g.*, corresponding to all of the expressed genes in a given cell type, are well known.

In practicing the method of the invention, the polynucleotides in the mixture are labeled with a fluorescent reporter. The fluorescent labeling can be carried out by standard methods, typically by transcribing DNA in the presence of fluorescence labeled nucleotides, such as Cy5-dNTP or fluorescent-labeled dNTP. Labeling may be carried out *in vitro* or *in vivo*.

In one embodiment of the invention, described below, mixture of polynucleotides from two different cell sources, referred to herein as test and control cells, are labeled with different fluorescent reporters which have independently detectable fluorescent properties, typically different fluorescent-emission peaks. The two independently labeled mixtures are then combined.

The mixture of labeled polynucleotides is contacted with the microarray surface, under conditions which permit hybridization between microarray sequences and complementary sequences in the mixture. The amount of polynucleotide mixture added to the microarray should be less than that which would saturate most of the array regions. In a preferred embodiment, a mixture of poly-

nucleotides having a solution concentration of between about 0.1 to 10 μg DNA/μl is placed on the microarray surface, in a total volume of between about 0.5-10 μl/cm^2 microarray surface, and the surface is covered with a glass coverslip. In this configuration, the "thickness" of the polynucleotide mixture is preferably between 10 and 500 μm.

When the polynucleotide mixture is double-stranded DNA, the double-stranded species may be denatured before application to the microarray surface, *e.g.*, by heating above the denaturation temperature, then rapidly cooling, as described in Example 1 below. Alternatively, the double-stranded material may be denatured by heating after addition to the slide, then cooled to permit hybridization.

Standard hybridization conditions for hybridizing labeled polynucleotides to the complementary-sequence DNA sequences on the microarray are employed, preferably to an end point at which complete hybridization is achieved, as detailed in the Methods below. The salt and temperature conditions of hybridization are selected for a desired stringency, according to well known principles, and/or the array is washed after hybridization at high stringency salt and temperature conditions.

After hybridization, the microarrays are washed, then scanned for fluorescence intensity. The scanning may be carried out with a confocal laser scanning device, as detailed in the Methods below. From the scanned fluorescence data, an image of microarray can be reconstructed, typically employing software to convert image intensity, which may vary over 3-4 logs, to a color or gray-scale intensity.

Where the array being scanned contains two fluorescent labels, each label is scanned independently at a selected excitation wavelength. After correcting for

optical crosstalk between the fluorophores, due to their overlapping emission spectra, the combined pattern may be represented as two separate images, in which the intensity of each label at each array position is represented, for example, in terms or a color or gray scale. Figs. 4A and 4B, and 5A and 5B, discussed below, are illustrative.

Alternatively, in the case of a two-color image, each label may be represented with a different color, *e.g.*, red and green, at a color intensity corresponding to the measured intensity of that fluorophore. This representation has the advantage of allowing quick visual assessment of microarray regions where one or the other fluorophore dominates in polynucleotide abundance (as evidenced, for example, by predominantly red or green spots), or where equal abundances of the two species are present (as evidenced, for example, by yellow spots). This type of representation is seen in Figs. 6-8, discussed below.

As discussed below, the method allows for detection in changes of relative abundance between different polynucleotides on the array, and between the same polynucleotide from two different sources, of 20-50% or less. Thus, where the method is used to determine relative levels of gene expression in a plurality of genes from two different cell sources, very small changes in expression, *e.g.*, 20%, can be detected. As will be illustrated below, this sensitivity allows large numbers of genes, *e.g.*, all genes whose mRNA is present at a level or at least $1:10^6$ and $1:10^7$, to be simultaneously monitored for both major and slight variations in gene expression, in response to a given change in cell state or cell type. This capability, in turn, allows for the simple identification of many new genes that are up-regulated or down-regulated in response to a particular shift in cell state or type.

IV. Applications

The method described above has a variety of applications, including genetic and physical mapping of genomes, genetic diagnosis, genotyping of organisms, monitoring of gene expression, and gene discovery.

A. Genetic Analysis

For genetic analysis, a plurality of polynucleotides, *e.g.*, cloned genomic fragments, is hybridized to an ordered array of DNA fragments--typically cloned genomic fragments-- and the identity of the DNA elements applied to the array is unambiguously established by the pixel or pattern of pixels of the array that are detected.

One application of such arrays for creating a genetic map is described by Nelson, *et al.* (1993). In constructing physical maps of the genome, arrays of immobilized cloned DNA fragments are hybridized with other cloned DNA fragments to establish whether the cloned fragments in the probe mixture overlap and are therefore contiguous to the immobilized clones on the array. For example, Lehrach, *et al.* (1990), describe such a process.

Example 1 illustrates an application of the method, for studying genomic complexity in *S. cerivisiae*. Here genomic fragments from the six largest yeast chromosomes were labeled with one fluorescent tag, and fragments from the 10 smallest chromosomes, with another tag, and the labeled two fragment mixtures were hybridized with a microarray of DNA sequences representing cloned yeast genomic fragments. The results are shown in Fig. 3A. A red signal in the figure indicates that the lambda clone on the array surface contains a cloned genomic DNA segment from one of the largest six yeast chromosomes. A green signal indicates that the lambda clone insert

comes from one of the smallest ten yeast chromosomes. Orange signals indicate repetitive sequences which cross hybridized to both chromosome pools. Control spots on the array confirm that the hybridization is specific and reproducible.

The physical map locations of the genomic DNA fragments contained in each of the clones used as array elements have been previously determined by Olson and co-workers (Riles, *et al.*, 1993), allowing for the automatic generation of the color karyotype shown in Figure 3B.

The color of a chromosomal section on the karyotype corresponds to the color of the array element containing the clone from that section. The black regions of the karyotype represent false negative dark spots on the array (10%) or regions of the genome not covered by the Olson clone library (90%). The largest six chromosomes are mainly red while the smallest ten chromosomes are mainly green, matching the original CHEF gel isolation of the hybridization probe. Areas of the red chromosomes containing green spots and vice-versa are probably due to spurious sample tracking errors in the formation of the original library and in the amplification and spotting procedures.

It can be appreciated how this approach can be applied to other types of genetic analysis, for example, in comparative genomic hybridization, or for use determining the extent of genetic divergence between two different species.

In all of these applications, the microarray method allows a large number of genomic sequences, *e.g.*, 10^4-10^5 or more, to be examined on a single array, using only small amount of genetic material, and with the capability of detecting changes or differences in the relative abundance between labeled polynucleotides on the array of as little as 20%.

B. Gene Expression Analysis

In another general application, a microarray of cDNA clones representing genes from a given source, or the expressed genomic sequences from that source, is hybridized with labeled cDNA's to monitor gene expression for research or diagnostic purposes. In a one-color mode, the method is used to examine relative expression levels among the various genes represented on the microarray. Alternatively, to monitor relative levels of gene expression between the individual genes from test and control cells, each labeled cDNA mixture can be hybridized with an individual microarray, and the patterns of fluorescence intensities in the two microarrays compared.

Relative levels of gene expression from two sources can be more advantageously studied, in accordance with the method, in single-array mode where the cDNA mixtures from two different cell sources are labeled with different fluorescent labels, and hybridized simultaneously with a single microarray. The ratio of the two fluorescent labels at each array position is a measure of the differential expression of that gene in the two cell sources. The ratio measurement is independent of the absolute abundance level of that gene.

It is also possible to use more than two fluorescent tags, when examining more than two cell sources simultaneously. This approach greatly enhances the resolution of the method, *i.e.*, the ability to detect small changes in expression level, by effectively eliminating variations between two separate microarrays in fluorescent signals due to, for example, different amounts of target sequences on the arrays.

The use of the two-color method for examining gene expression levels in plants in illustrated in Example 2. Briefly, mRNA obtained from wildtype and transgenic *arabidopsis* plants containing an exogenous HAT-4 gene

were isolated, and cDNA from the two sources was prepared and labeled with different fluorescent labels. The cDNA mixture was hybridized to a single microarray containing an array of *arabidopsis* cloned cDNA's. Figs. 4A and 4B show the fluorescent-intensity scans of the microarray for the wildtype cDNA fluorescence label (4A) and the transgenic cDNA fluorescence label (4B) (both from the same microarray). The gene expression patterns differ in several respects, but in particular, by the presence in two strong spots in Fig. 4B (indicated at 49.50) corresponding to HAT-4 cDNA.

Using an identical same microarray, a similar study was carried out to examine differences in the gene expression patterns from root and leaf tissue from *arabidopsis*, shown in Figs. 5A and 5B, respectively. As seen, the two tissues give quite different patterns of expression for the particular cDNA sequences on the array. Details are given in Example 2.

One of the important applications of the method is the ability to identify genes that are associated with a given cell state or type. The objective of such studies is to identify new genes that can be used as diagnostic indicators of the cell state or type, and to identify new cellular enzymes and pathways, related to expression of the newly discovered genes, that can serve as targets in developing new therapeutic agents.

Heretofore, detecting levels of differential gene expression between control and test cells has been limited by the number of genes that can be examined on a single array, and the sensitivity of array methods to changes in the levels of hybridized probe. In general, these limitations significantly increase the work required to identify low-copy-number genes or interest, and may even prevent such genes from being identified.

To illustrate the method as applied to identifying new genes in human bone-stem cells, a microarray of human cDNA clones, picked at random from a human peripheral blood lymphocyte cDNA library, was prepared as detailed in Example 3. The array contained 1,046 PBL clones, and 10 sequences from *arabidopsis*, as control sequences.

In a first study, a mixture of cDNA's from bone marrow labeled with Cy5-dCTP and cDNA's from Jurkat cells labeled with fluorescein-dCTP were hybridized to the microarray, to assess the different patterns of gene expression in the two blood-cell types. The scanned array is shown in Fig. 6, where red spots indicate higher gene expression in bone marrow, green spots, greater gene expression levels in Jurkat cells, and yellow spots, comparable gene expression levels in both cells. It is apparent from the figure that the two cell types have quite different levels of gene expression in many genes.

A second study was designed to identify genes responsive to heat shock in the Jurkat cell line. A number of heat-shock proteins in Jurkat and other cell lines have been identified, as discussed below. It was therefore of interest to see if the present method could extend the number of genes known to be responsive.

In the study, Jurkat cells, after initial culturing, were grown for 4 hours at 37°C and 43°C, respectively. Total mRNA from the harvested control and heat-shocked cells was labeled by reverse transcriptase incorporation of fluorescein- or Cy5-derivatized dCTP, respectively, as detailed in Example 3. The two cDNA fractions were mixed and hybridized to the human PBL microarray above. The combined fluorescent scans for the two fluorophores are shown in Fig. 7.

Examination of the fluorescent scans revealed positive hybridization signals to >95% of the human cDNA array elements, but not to any of the Arabidopsis con-

trols. Hybridization intensities spanned more than three orders of magnitude for the 1,046 array elements surveyed. Comparative expression analysis of heat shocked versus control cells in the two experiments revealed altered fluorescence intensities at 17 array elements. Of the 17 putative differentially expressed genes, 11 were induced by heat shock treatment and 6 displayed modest repression. This result is indicated schematically in Fig. 9A which shows up-regulated and down-regulated genes as red and green spots, respectively.

To determine the identity of the genes that exhibited altered expression patterns, cDNAs corresponding to each of the 17 array elements were subjected to single pass DNA sequencing on the proximal end of each clone (see Example 3 for details). Database searches of the sequences revealed "hits" for 14 of the 17 clones (Table 1, B1-B17); the three remaining clones (B8-B10) did not match any sequence in the public human database, though one of the clones (B7) exhibited significant homology to an expressed sequence tag (EST) from *C. elegans*. To further confirm the identity of the clones, the nucleotide sequence of distal end each cDNA was determined. In all cases, proximal and distal cDNA sequences mapped to the same gene.

Table 1

Clone	Row	Column	Ratio	Blast Identity	Accession #
B1	2421	0.5	--	CYC oxidase III	J01415, J01415
B2	1	31	0.5	β-Actin	N.R., X00351
B3	15	8	0.5	CYC oxidase III	J01415, J01415
B4	32	19	0.5	CYC oxidase III	J01415, J01415
B5	17	8	0.5	CYC oxidase III	J01415, J01415
B6	22	31	0.5	β-Actin	X.R., X00351
B7	5	4	2.0	Novel*	U56653, U56654
B8	2	19	2.0	Novel*	U56655, U56656
B9	14	5	2.2	Novel*	U56657, U56658
B10	7	8	2.4	Polyubiquitin	X04803, X04803
B11	12	2	2.4	TCP-1	X52882, X52882
B12	28	2	2.5	Polyubiquitin	M17597, M17597
B13	14	7	2.5	Polyubiquitin	X04803, X04803
B14	20	9	2.6	HSP90°	M16660, M16660
B15	30	12	4.0	DnaJ homolog	D13388, D13388
B16	10	5	5.8	HSP90a	X07270, X07270
B17	13	16	6.3	HSP90a	M27024, X15183
B18	7	19	2.0	°-2-microglobulin	S54761, M30683
B19	21	30	2.1	Novel*	U56659, U56660
B20	3	26	2.2	°-2-microglobulin	S54761, M30683
B21	1	18	2.6	PPG kinase	M11968, L00160
B22	22	30	3.5	NF-kB1	Z47744, M55643
B23	20	16	19	PAC-1	L11329, L11329

The five most highly induced genes in heat-treated cells included heat shock protein 90a (HSP90α), DnaJ, HSP90β, polyubiquitin, and t-complex polypeptide-1 (TCP-1) (Table 1). HSPα, DnaJ, and HSP90β exhibited a

6.3-, 4.0-, and 2.6-fold induction, respectively; lesser activation was observed for genes encoding polyubiquitin and TCP-1 (Table 1). Three novel sequences (B7-B9) each exhibited induction in the 2-fold range (Table 1). A modest repression was observed for both β-actin and cytochrome c (cyc) oxidase III (Table 1). In several cases, clones corresponding to the same gene were recovered at multiple locations on the array; expression ratios for these clones varied by less than 10% from element to element (Table 1).

To confirm that the changes in expression determined with the microarray assay corresponded to altered mRNA levels, each of the cloned sequences were used as probes in RNA blotting analyses, as described in Example 3. All of the genes that displayed heat shock induction by microarray analysis, yielded similar results in "dot blot" experiments (Figs. 10A-10L and (Table 2, B1-B17). The gene encoding HSP90α, for example, exhibited 6.3-fold activation by microarray analysis and 7.2-fold induction by RNA blotting (Fig. 10I; Table 2). In all cases, expression ratios as determined by the two procedures differed by less than 2-fold for the genes identified in the heat shock experiments (Table 2). The two assays differed more widely in terms of assessing absolute expression levels; nonetheless, absolute expression as monitored on a microarray typically correlated with RNA blots to within a factor of five (Table 2).

Table 2

Clone	Blast Identity	Expression Level (per 105)			
		Microarray	Ratio	RNA Blot	Ratio
B1	CYC oxidase III	92/46	0.5	100/80	0.8
B2	β-Actin	240/120	0.4	270/280	1.0
B3	CYC oxidase III	36/18	0.5	N.D.	N.D.
B4	CYC oxidase III	76/38	0.5	N.D.	N.D.
B5	CYC oxidase III	62/31	0.5	N.D.	N.D.
B6	°-Actin	180/89	0.5	N.D.	N.D.
B7	Novel (weakly to D76026)	1.3/2.6	2.0	0.77/1.8	2.3
B8	Novel	2.0/4.0	2.0	1.5/3.4	2.3
B9	Novel	0.8/1.8	2.2	1.2/1.8	1.5
B10	Polyubiquitin	0.8/72	2.4	25/89	3.6
B11	TCP-1	2.3/77	2.4	7.1/27	3.8
B12	Polyubiquitin	0.8/2.0	2.5	N.D.	N.D.
B13	Polyubiquitin	1.7/4.3	2.5	N.D.	N.D.
B14	HSP90°	75/200	2.6	30/120	4.0
B15	DnaJ homolog	1.0/4.0	4.0	1.6/13	8.1
B16	HSP90a	0.6/3.5	5.8	3.2/29	9.1
B17	HSP90a	0.8/5.0	6.3	8.6/6.2	7.2

Genes that exhibited positive regulation in heat-treated T cells encode factors that either function as molecular "chaperones" of protein folding (HSP90α, HSP90β, DnaJ, TCP-1), or as mediators of selective protein degradation (polyubiquitin). The identification of these sequences is consistent the biochemical basis of heat shock induction. Many proteins undergo denaturation at elevated temperatures, and those that fail to maintain

proper conformation must be selectively degraded (Jindal, 1996; Wilkinson, 1995; Jakob, *et al.*, 1994; Becker and Craig, 1994; Cyr, 1994; Craig, *et al.*, 1994). It will be interesting to determine whether the three novel heat shock-inducible sequences (B7-B9) identified in this study play a role in protein folding and turnover, or possess some other biochemical activity. Complete nucleotide sequence determination, conceptual translation, expression monitoring, and biochemical analysis should provide a clue to the function of these genes.

In summary, the method was successful in confirming a number of known heat-shock genes and identifying three new genes not previously associated with heat shock.

A third study was designed to identify genes whose level of expression is related to treatment with phorbol ester. Phorbol ester, a potent activator the protein kinase C family (Newton, 1995; Nishizuka, 1995), activates a set of genes distinct from those involved in the heat shock pathway.

Control and phorbol-ester treated Jurkat cells were cultured as described in Example 4. Total cDNA from control and drug-treated cells were labeled by reverse transcriptase incorporation of fluorescein- or Cy5-derivatized dCTP, respectively, as detailed in Example 4. The two cDNA fractions were mixed and hybridized to the human PBL microarray above. The combined fluorescent scans for the two fluorophores are shown in Fig. 8.

As above, examination of the fluorescent scans revealed positive hybridization signals to >95% of the human cDNA array elements, but not to any of the Arabidopsis controls. Hybridization intensities spanned more than three orders of magnitude for the 1,046 array elements surveyed. Comparative expression analysis of heat shocked versus control cells in the two experiments

revealed altered fluorescence intensities at 6 array elements (Table 1 above, B18-B23), all of which showed modest to strong up-regulation in response to phorbol-ester treatment. The positions of the genes in the array are illustrated in Fig. 9B (red spots).

To determine the identity of the genes that exhibited altered expression patterns, cDNAs corresponding to each of the 6 array elements identified above were subjected to single pass DNA sequencing on the proximal end of each clone, as above. Database searches of the sequences revealed "hits" for 5 of the 6 clones (Table 1, B18-B23), including the two most highly induced genes (Table 1) which corresponded to a tyrosine phosphatase (PAC-1) (Rohan, *et al.*, 1993) and nuclear factor-kappa B1 (NF-kB1) (Thanos, *et al.*, 1995; Baeuerle, *et al.*, 1994; Liou, *et al.*, 1993). Modest activation was also observed for genes encoding phosphoglycerate kinase (PGK), β-2-microglobulin, and one additional sequence (B19) that did not match any entry in the public database (Table 1).

Each of the phorbol ester-inducible genes identified by microarray analysis displayed increases in steady-state mRNA levels (Figs. 11A-11E; Table 2, B18-23) in RNA blotting experiments.

It is striking that, despite the previous intensive analyses of both the heat shock and phorbol ester pathways, 4 of the 15 sequences identified in the two studies above represent novel human genes. The fact that the four novel genes share the common features of relatively low expression (about 1:50,000) and modest activation (about 2.2-fold), suggest that these sequences may have been simply overlooked in screens utilizing prior-art differential techniques.

One way to examine the function of newly discovered genes is to determine their expression profiles in different tissues. To explore this, probes were prepared

from human bone marrow, brain, prostate and heart by labeling mRNA with reverse transcriptase in the presence of Cy-5-dCTP. In a separate reaction, a control probe was prepared by labeling total Jurkat mRNA with fluorescein-dCTP. The four Cy-5-labeled tissue samples were each mixed with an aliquot of the fluorescein-labeled Jurkat probe, and the two-color probe mixtures were hybridized to four separate microarrays. The four arrays were then washed and scanned for fluorescence emission. Hybridization signals for each of the tissues samples were normalized internally to the Jurkat control and an expression profile was generated for each of the 1,046 genes represented on the array.

Detectable expression was observed for all 15 of the heat shock- and phorbol ester-regulated genes in the four tissue types examined. In general, the expression level of each gene in Jurkat cells correlated rather closely with expression in the four tissues. Genes encoding β-actin and cytochrome c oxidase, the two most highly expressed of the 15 genes in Jurkat cells (Table 2), were also highly expressed in bone marrow, brain, prostate, and heart (Fig. 12A); similarly, genes expressed at moderate or low levels in Jurkat cells (Table 2) displayed moderate to weak expression in the four tissue types (Figs. 12B, 12C). One the novel heat shock genes (B7) showed an expression profile similar to the mitochondrial gene encoding CYC oxidase III.

From the foregoing, it can be appreciated how various objects and features of the method of the invention are met. The microarrays used in the experiment can be designed for probing a wide range of genes, *e.g.*, from particular species, cell types, or cell states, and an array density which allows large numbers of genes, *e.g.*, 10^5-10^6 to be examined on a single array surface. This not only reduces the amount of work needed to screen

large numbers of genes, but because a single array provides its own internal controls, allows higher resolution in determining relative levels of abundance of polynucleotide probe species on the array.

In the two-color, single-array mode, the method offers the additional advantage of an internal control in the levels of fluorescent signals associated with cDNA's obtained from two different sources. The greater resolution this features provides allows detection of differential gene expression, either up-regulation or down-regulation, of as low as 20%. As a result, it is possible not only to scan very large collections of gene transcripts on a single array, but also to detect relatively minor changes in gene expression in each array sequence. Another advantage of the two color scheme is that even if the number of binding sites of the arrayed target DNA elements becomes limiting, the competitive hybridization of the two samples saturate the sites in a ratio reflecting their relative abundance.

The studies involving genes responses to heatshock or phorbol-ester treatment illustrate the ability of the method to rapidly identify new genes associated with such cell states.

The following examples illustrate, but in no way are intended to limit, the present invention.

V. General Methods

A. Microarray Preparation

Target messenger nucleic acid DNA fragments were prepared as described in the examples. In one general embodiment (DNA sequences non-covalently linked to microarray surface), the microarrays were fabricated on microscope slides which were coated with a layer of poly-l-lysine (Sigma). For each microarray region, an automated apparatus loaded 1 μl of the concentrated target

DNA in 3 × SSC directly from 96 well storage plates into the open capillary printing element and deposited about 5 nl of sample per slide at desired spacing between spots (Shalon, 1995), as described also in above-cited WO95/35505.

After the spotting operation was complete, the slides were rehydrated in a humid chamber for 2 hours, baked in a dry 80° vacuum oven for 2 hours, rinsed to remove un-absorbed DNA and then treated with succinic anhydride to reduce non-specific adsorption of the labeled hybridization probe to the poly-l-lysine coated glass surface. Immediately prior to use, the immobilized DNA on the array was denatured in distilled water at 90° for 2 minutes.

In another general embodiment (DNA sequences co-valently linked to the microarray surface), target sequences containing a reactive primary amine group on their 5′ end were arrayed on a 1.0 cm^2 glass surface of a silylated microscope slides (CEL Associates), using the high-speed robotic printing method described above.

The target DNA sequences were linked covalently to the glass surface and heat denatured to allow hybridization. Briefly, the printed arrays were incubated for 4 hours in a humid chamber to allow rehydration of the array elements, rinsed once in 0.2% sodium dodecyl sulfate (SDS) for 1 min, twice in H20 for 1 min, and once in sodium borohydride solution (1.0 g NaBH4 dissolved in 300 ml phosphate buffered saline (PBS) and 100 ml 100% ethanol). The arrays were submerged in H20 for 2 min at 95°C, transferred quickly into 0.2% SDS for 1 min, rinsed twice in H20, air dried, and stored in the dark at 25°C.

B. Preparation of Reporter-Labeled Messenger Nucleic Acid

Total RNA was isolated from a selected cell or tissue source using standard methods (Sambrook, *et al.*, 1988). PolyA+ mRNA was prepared from total RNA using "OLIGOTEX-DT" resin (Qiagen). Reverse transcription reactions were carried out using a "STRATASCRIPT" RT-PCR kit (Stratagene) modified as follows: 50 μl reactions contained 0.1 μg/μl mRNA, 0.1 ng/μl human acetylcholine receptor mRNA, 0.05 μg/μl oligo-dT (21mer), 1X first strand buffer, 0.03 units/μl RNase block, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM fluorescein-12-dCTP (or lissamine-5-dCTP) and 0.03 units/μl "STRATASCRIPT" reverse transcriptase. Reactions were incubated for 60 min at 37°C, precipitated with ethanol, and resuspended in 10 μl TE.

The samples were then heated for 3 min at 94°C and chilled on ice. RNA was degraded by adding 0.25 μl 10N NaOH followed by a 10 min incubation at 37°C. The samples were neutralized by adding 2.5 μl 1M Tris-HCl (pH 8.0) and 0.25 μl 10N HCl, and precipitated with ethanol. Pellets were washed with 70% ethanol, dried to completion in a "SPEEDVAC" (Savant, Farmingdale, NY) resuspended in 10 μl H_2O, and reduced to 3.0 μl in a speedvac. Fluorescent nucleotide analogs were purchase from DuPont NEN (Boston, MA).

C. Hybridization of Reporter-Labeled Nucleic Acid to Target DNA

Hybridization reactions contained 1.0 μl of fluorescent cDNA synthesis product (~2 μg) and 1.0 μl of hybridization buffer (10 × SSC, 0.2% sodium dodecyl sulfate; SDS). The 2.0 μl probe mixtures were aliquoted onto the microarray surface and covered with 12 mm round cover slips. Arrays were transferred to a waterproof slide

chamber having a cavity just slightly larger than a microscope slide. The chamber was kept at 100% humidity internally by the addition of 2 microliters of water in a corner of the chamber. The chamber containing the arrays was incubated for 18 hr at 65°C.

The arrays were washed for 5 min at room temperature (25°C) in low stringency wash buffer (1 × SSC, 0.1% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 × SSC, 0.1% SDS).

D. Detection of Hybridized Sequences

The microscope used to detect the reporter-labeled hybridization complexes was outfitted with an Innova 70 mixed gas 10 W laser (Coherent Lasers, Santa Clara, CA) capable of generating a number of spectral lines, including lines at 488 nm and 568 nm, and 632 nm (for Cy5). The excitation laser light was focused on the array using a 20X microscope objective (Nikon).

The slide containing the array was placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm × 1.8 cm array used in the present example was scanned with a resolution of 20 μm. Spatial resolutions up to a few micrometers are possible with appropriate optics.

In two separate scans, a mixed gas multiline laser excited the two fluorophores sequentially. Emitted light was split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics, San Jose, CA) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes were used to filter the signals. The emission maxima of the fluorophores used were 517 nm (fluorescein), 588 nm (lissamine), and 650 for Cy5. Each array was typically scanned twice -- one scan per fluorophore, using the appropriate filters at the laser source

-- although the apparatus was capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans was typically calibrated using the signal intensity generated by an mRNA or cDNA control species added to the hybridization mix at a known concentration. For example, in the experiments described in Example 3, human acetylcholine receptor mRNA was added to the wild-type *Arabidopsis* poly-A total mRNA sample at a weight ratio of 1:10,000. A specific location on the array contained a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:10,000.

When messenger nucleic acids derived probes containing two different fluorophores (*e.g.*, representing test and control cells) are hybridized to a single array for the purpose of identifying genes that are differentially expressed, a similar calibration scheme may be employed to normalize the sensitivity of the photomultiplier tubes such that genes expressed at the same levels in the test and control samples display the same pseudocolor intensity. In one embodiment, this calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

It will be understood that where greater confidence in the absolute levels of expression is desired, multipoint calibrations may be performed.

E. Analysis of Patterns of Reporter Levels

The output of the photomultiplier tube was digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood, MA) installed in an IBM-compatible PC computer. The digitized data were displayed as an image where the signal intensity was

mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal).

The data were also analyzed quantitatively. In cases where two different fluorophores were excited and measured simultaneously, the data were first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid was superimposed over the fluorescence signal image such that the signal from each spot was centered in each element of the grid. The fluorescence signal within each element was then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for the above analyses was similar in functionality to "IMAGE-QUANT", available from Molecular Dynamics (Sunnyvale, CA).

Example 1

Genomic-Complexity Hybridization to Micro DNA Arrays Representing the Yeast *Saccharomyces cerevisiae* Genome with Two-Color Fluorescent Detection

The array elements were randomly amplified PCR (Bohlander, *et al.*, 1992) products using physically mapped lambda clones of *S. cerevisiae* genomic DNA templates (Riles, *et al.*, 1993). The PCR was performed directly on the lambda phage lysates resulting in an amplification of both the 35 kb lambda vector and the 5-15 kb yeast insert sequences in the form of a uniform distribution of PCR product between 250-1500 base pairs in length. The PCR product was purified using Sephadex G50 gel filtration (Pharmacia, Piscataway, NJ) and concentrated by evaporation to dryness at room temperature overnight. Each of the 864 amplified lambda clones

was rehydrated in 15 μl of 3 × SSC in preparation for spotting onto the glass.

The microarrays were fabricated on microscope slides coated with a layer of poly-l-lysine, as above. Immediately prior to use, the immobilized DNA on the array was denatured in distilled water at 90° for 2 minutes.

For the pooled chromosome experiment, the 16 chromosomes of *Saccharomyces cerevisiae* were separated in a CHEF agarose gel apparatus (Biorad, Richmond, CA). The six largest chromosomes were isolated in one gel slice and the smallest 10 chromosomes in a second gel slice. The DNA was recovered using a gel extraction kit (Qiagen, Chatsworth, CA). The two chromosome pools were randomly amplified in a manner similar to that used for the target lambda clones. Following amplification, 5 micrograms of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham, Arlington Heights, IL) with a lissamine conjugated nucleotide analog (Dupont NEN, Boston, MA) for the pool containing the six largest chromosomes, and with a fluorescein conjugated nucleotide analog (BMB) for the pool containing smallest ten chromosomes. The two pools were mixed and concentrated using an ultrafiltration device (Amicon, Danvers, MA).

Five micrograms of the hybridization probe consisting of both chromosome pools in 7.5 μl of TE was denatured in a boiling water bath and then snap cooled on ice. 2.5 μl of concentrated hybridization solution was added (final concentration 5 × SSC and 0.1% SDS), and all 10 μl transferred to the array surface, covered with a cover slip, placed in a custom-built single-slide humidity chamber and incubated at 60° for 12 hours. The slides were then rinsed at room temperature in 0.1 × SSC and 0.1%SDS for 5 minutes, cover slipped and scanned.

After correcting for optical crosstalk between the fluorophores due to their overlapping emission spectra, the red and green hybridization values for each clone on the array were correlated to the known physical map position of the clone resulting in a computer-generated color karyotype of the yeast genome, as shown in Figs. 3A and 3B, discussed above.

Example 2

Fluorescence Detection of Gene Expression Patterns using Micro Arrays of Arabidopsis cDNA Clones

A. Microarray Preparation

Target messenger nucleic acid DNA fragments were made by amplifying the gene inserts from 45 different *Arabidopsis thaliana* cDNA clones and 3 control genes using the polymerase chain reaction (PCR; Mullis, *et al.*). The DNA fragments comprising the PCR product from each of the 48 reactions were purified using "QIAQUICK" PCR purification kits (Qiagen, Chatsworth, CA), eluted in ddH_2O, dried to completion in a vacuum centrifuge and resuspended in 15 μl of 3X sodium chloride/sodium citrate buffer (SSC). The capacity of the "QIAQUICK" purification kits is 10 μg of DNA; accordingly, each sample contained about 10 μg or less of DNA.

The samples were then deposited in individual wells of a 96 well storage plate with each sample split among two adjacent wells as a test of the reproducibility of the arraying and hybridization process. The samples were spotted on poly-l-lysine-coated microscope slides, as above to produce a microarray with regions about 500 μm apart.

The positions of several specific elements in the 96-element array, and the reasons for their inclusion, are indicated in Table 3, below. The remaining elements

of the array consist of known or unknown genes selected from an *Arabidopsis* cDNA library.

Table 3

Element #	Name	Purpose
1, 2	Human acetylcholine receptor gene	Control for expression level
13, 14	Chlorophyll binding protein gene	Gene with known expression
35, 36	Rat glucocorticoid receptor gene	Positive and negative control
49, 50	HAT4 transcription factor gene	Gene with known expression
95, 96	Yeast TRP4 gene	Positive and negative control

B. Methods

Total polyA+ RNA was isolated from plant tissue of *Arabidopsis* using standard methods, and was used to prepare cDNA labeled with either fluorescein-12-dCTP or lissamine-5-dCTP, as above. Hybridization and scanning were carried out as above.

C. Two-Color Detection of Differential Gene Expression in Wild Type versus Transgenic *Arabidopsis* Tissue

Differential gene expression was investigated using a simultaneous, two-color hybridization scheme, which served to minimize experimental variation inherent in comparing independent hybridizations. Two μg of wild-type *Arabidopsis* total cDNA that were labeled with fluorescein (as above) were combined with two micrograms of transgenic *Arabidopsis* total cDNA that were labeled by incorporating lissamine-5-dCTP (DuPont NEN) in the reverse transcription step and hybridized simultaneously

to a microarray containing the same pattern of spotted cDNAs as described above.

To test whether overexpression of a single gene could be detected in a pool of total *Arabidopsis* mRNA, methods of the invention were used to analyze a transgenic line overexpressing the transcription factor *HAT4* (Schena, *et al.*, 1995). The transgenic *Arabidopsis* tissue was known to express HAT4 at levels of 0.5% of the total transcripts, while wild-type expression of HAT4 was only 0.01% of total transcripts (as previously determined by Northern analysis; Schena, *et al.*, 1995).

Human acetylcholine receptor mRNA was added to the wild-type *Arabidopsis* poly-A total mRNA sample at a weight ratio of 1:10,000 and into the transgenic *Arabidopsis* poly-A total mRNA sample at a weight ratio of 1:100 to roughly match the expected expression levels of HAT4.

As a cross-check of the negative controls, linear PCR was used to generate single-stranded fluorescein-labeled rat glucocorticoid receptor DNA and lissamine-labeled yeast TRP4 DNA. The two PCR products were added to the hybridization solution at a partial concentration of ~1:100. The two fluorophores were excited separately in two separate scans in order to minimize optical crosstalk.

The array was then scanned separately for fluorescein and lissamine emission following independent excitation of the two fluorophores as described in Example 2, above. The results of the experiments are shown in Fig. 4A and 4B.

D. Two-Color Detection of Differential Gene Expression in Root versus Leaf Tissue

In a similar experiment using the same labeling and hybridization procedures described above, 2 μg of total

cDNA from *Arabidopsis* root tissue labeled with fluorescein were combined with two micrograms of total cDNA from *Arabidopsis* leaf tissue labeled with lissamine and were simultaneously hybridized to a micro array containing the same pattern of target sequences described above. The acetylcholine gene mRNA was added to both poly-A total mRNA samples at 1:1,000 to allow for normalization of fluorescence intensities. The glucocorticoid and TRP4 controls were added to the hybridization probe as before. The results are shown in Figs. 5A and 5B.

Example 3
Differential Gene Expression Due to Heat Shock in Human T cells (Jurkat Cell Line)

A. Constructing a Human Gene Expression Microarray

Human cDNA clones were picked at random from a human peripheral blood cDNA library, and propagated as bacterial cultures. The human cDNA library was made using mRNA isolated from human peripheral lymphocytes transformed with the Epstein-Barr Virus (EBV). Inserts >600 bases were cloned into the lambda vector lYES-R to generate 107-108 recombinants. Bacterial transformants were obtained by infecting E. coli strain JM107/lKC. Colonies were picked, propagated in a 96-well format, and minilysate DNA was prepared by alkaline lysis using REAL preps (Qiagen).

Plasmid DNA was isolated and inserts from each clone were amplified by use of the polymerase chain reaction (PCR) and purified. Inserts were amplified by PCR in a 96-well format using primers (PAN132, 5′ CCTCTATACTTTAACGTCAAGG; PAN133, 5′ TTGTGTGGAATTGTGAGCGG) complementary to the lYES polylinker and containing a six carbon amino modification (Glen Research) on the 5′ end. PCR products were purified in a 96-well format using QIAquick columns (Qiagen).

A total of 1,056 purified PCR inserts representing 1,046 human clones and 10 Arabidopsis controls were arrayed onto a 1.0 cm^2 glass surface of a glass slide, and attached covalently to the slide surface as detailed above.

B. Examining the Heat Shock Response in Human Cells

Human T (Jurkat) cells were grown in a tissue culture incubator (37°C and 5% CO2) in RMPI medium supplemented with 10% fetal bovine serum, 100 μg/ml streptomycin, and 500 U/ml penicillin. The cells were propagated to near confluence under normal growth conditions, divided into two equal aliquots, and grown for 4 hours at 37°C and 43°C, respectively. Cells from the control (37°C) and heat shocked culture (43°C) were harvested, lysed (Ausubel, *et al.*, 1994), and total mRNA from the two cell samples was labeled by reverse transcriptase incorporation of fluorescein- or Cy5-derivatized dCTP, respectively.

Arabidopsis control mRNAs were made by in vitro transcription of cloned HAT4, HAT22, and YesAt-23 cDNAs (Cohen, *et al.*, 1995; Chan, *et al.*, 1994; Crabtree, *et al.*, 1994) using an RNA Transcription Kit (Stratagene). For quantitation, the mRNAs were doped into the RT reaction at ratios of 1:100,000, 1:10,000, and 1,1,000 (w/w) respectively.

Hybridizations and array scanning were carried out as above. To avoid complications arising from fluor-specific effects, the flours were "swapped" in a second set of labeling reactions, such that samples from control and heat shock-treated samples were labeled with Cy5- and fluorescein-dCTP, respectively. Each pair of fluorescent probes were mixed and hybridized to two 1,056 element human gene expression microarrays. The arrays

were washed at high-stringency and scanned with a confocal laser scanning device to detect emission of the two flours.

Examination of the fluorescent scans revealed positive hybridization signals to >95% of the human cDNA array elements, but not to any of the Arabidopsis controls. Hybridization intensities spanned more than three orders of magnitude for the 1,046 array elements surveyed. Comparative expression analysis of heat shocked versus control cells in the two experiments revealed altered fluorescence intensities at 17 array elements. Of the 17 putative differentially expressed genes, 11 were induced by heat shock treatment and 6 displayed modest repression.

C. Identification of Heat-Shock Related Genes

Sequencing reactions were carried out using the PAN132 and 133 primers and a 373A automated sequencer according to the instructions of the manufacturer (Applied Biosystems). Sequence searches were made to the non-redundant nucleotide database at the National Center for Biotechnology Information (NCBI) using Macintosh Blast software.

For dot-blot studies, samples of polyA+ mRNA (9) corresponding to 1.0, 0.1 and 0.01 μg, respectively, were suspended in 10 × SSC, spotted onto nylon membranes (Nytran), and crosslinked with ultraviolet light using a Stratalinker 1800 (Stratagene). Probes were prepared from cloned sequences (Table 1) by random priming using a Prime-It II kit (Stratagene) in the presence of P32-dATP. Hybridizations were carried out following the instructions of the manufacturer. Quantitation was performed on a PhosphorImager (Molecular Dynamics).

Example 4

Differential Gene Expression Due to Phorbol-Ester in Human T cells (Jurkat Cell Line)

To explore a signaling pathway distinct from the heat shock response, microarrays constructed as in Example 3 were used to examine the cellular effects of phorbol ester treatment. Jurkat cells were grown to near confluence, treated with phorbol ester, harvested, lysed and used source of mRNA, as above. Samples of mRNA from untreated or phorbol ester-stimulated cells were labeled by reverse transcriptase incorporation of fluorescent dCTP analogs, as above. The two-color fluorescent probes were mixed, hybridized to microarrays, and scanned for fluorescence emission. A total of six array elements displayed elevated fluorescent signals with probes derived from phorbol ester-treated cells relative to a controls (Fig. 8).

Although the invention has been described with respect to specific embodiments and methods, it will be clear that various changes and modification may be made without departing from the invention.

IT IS CLAIMED:

1. A method of determining the relative amounts of individual polynucleotides in a mixture of different-sequence polynucleotides, comprising

labeling the polynucleotides with a fluorescent reporter, to form a mixture of labeled polynucleotides,

contacting the labeled polynucleotides, under hybridization conditions, with a microarray of different DNA sequences disposed at discrete locations a non-porous surface, at a density of at least about 100 locations/cm^2, where the different DNA sequences in the array are each (i) present in multiple copies, and (ii) effective to hybridize to individual polynucleotides in the mixture, and

determining the level of fluorescence at each position in the microarray.

2. The method of claim 1, wherein said contacting includes covering the array surface with a solution of the mixture of labeled polynucleotides, to a solution depth of less than 500 microns.

3. The method of claim 1, wherein the DNA sequences in the array are at least about 50 bases in length.

4. The method of claim 1, wherein the labeled polynucleotides represent at least 1 million unique base sequences.

5. The method of claim 1, wherein the density of array elements corresponding to different-sequence DNA locations in the array is at least 1,000/cm^2.

6. The method of claim 1, for use in determining the relative amounts of each polynucleotide from first and second different sources, wherein (i) the polynucleotides from the first and second sources are labeled with independently detectable first and second fluorescent reporters, respectively, (ii) said contacting of labeled polynucleotides from first and second sources is carried out simultaneously under competitive hybridization conditions, and (iii) said determining includes measuring the levels of the two reporters at each position in the array.

ABSTRACT

A method of determining the relative amounts of individual polynucleotides in a complex mixture of different-sequence polynucleotides is disclosed. The polynucleotides, after fluorescent labeling, are contacted under hybridization conditions with an array of different DNA sequences disposed at discrete locations on a non-porous surface, at an array density of at least about 100 sequences/cm^2, where the different DNA sequences in the array are effective to hybridize to individual polynucleotides in the mixture. The level of fluorescence associated with each array sequence provides a measure of its relative amount in the mixture.

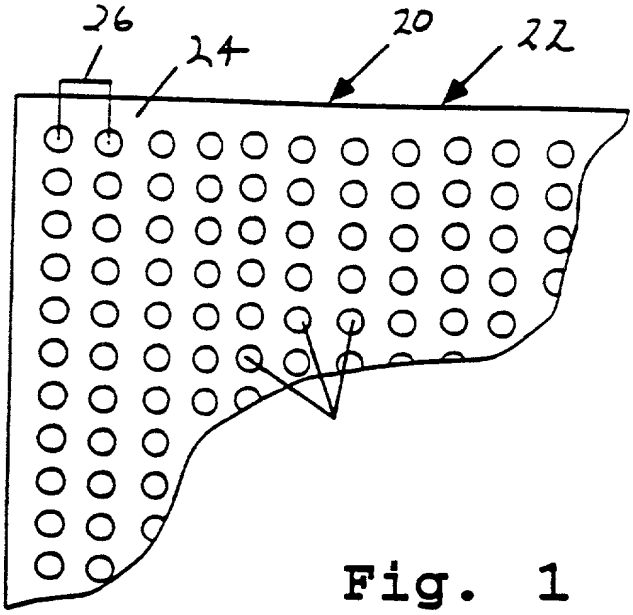

Fig. 1

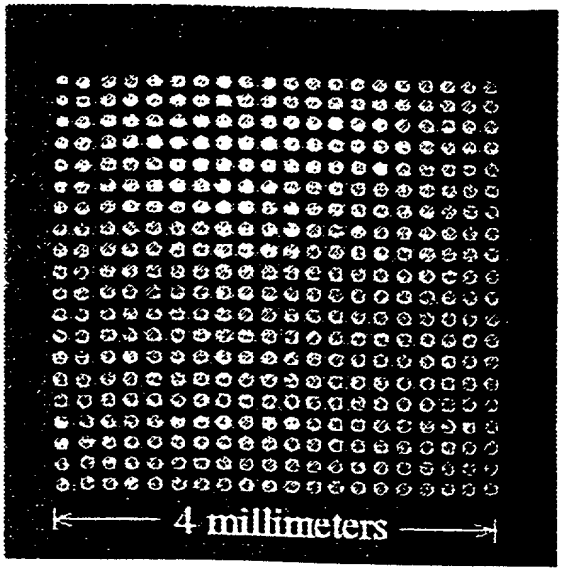

Fig. 2

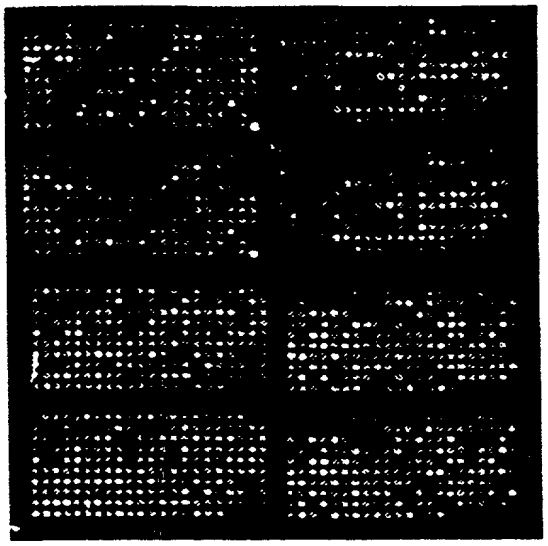

Fig. 3A

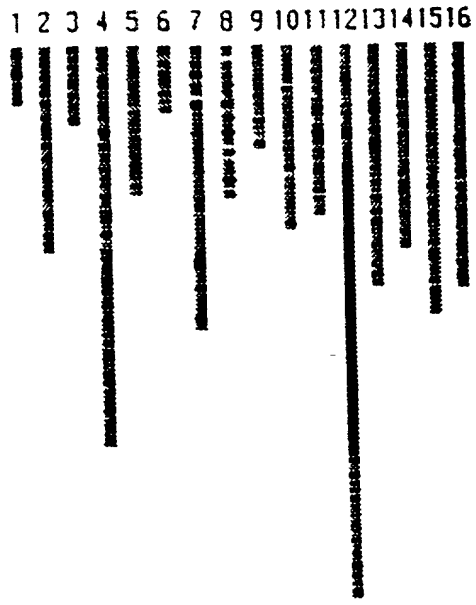

Fig. 3B

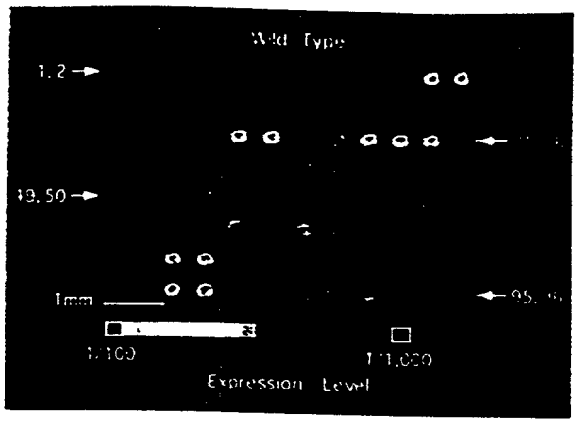
Wild Type
1.2
49.50
1mm
1/100
1/1,000
Expression Level

Fig. 4A

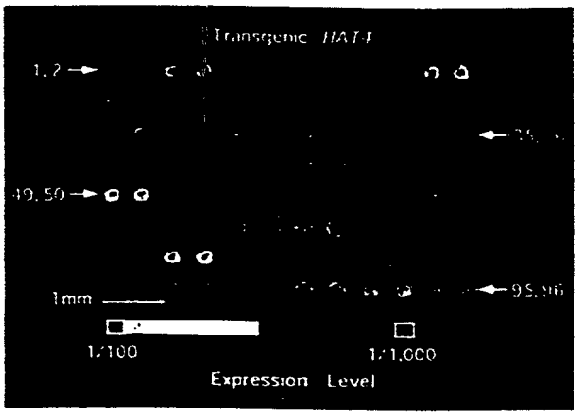
Transgenic HAT-1
1.2
49.50
1mm
1/100
1/1,000
Expression Level

Fig. 4B

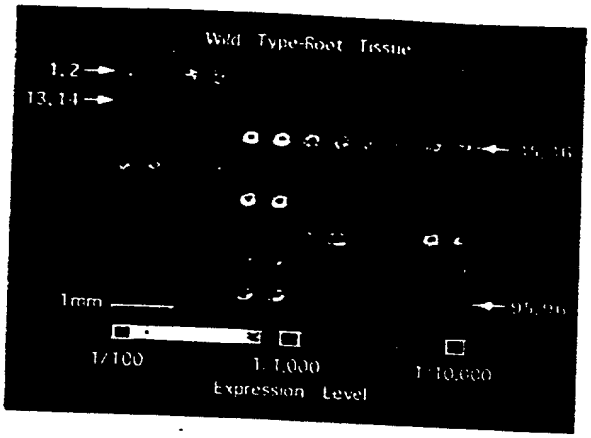
Wild Type-Root Tissue
1.2
13.14
1mm
1/100
1/1,000
1/10,000
Expression Level

Fig. 5A

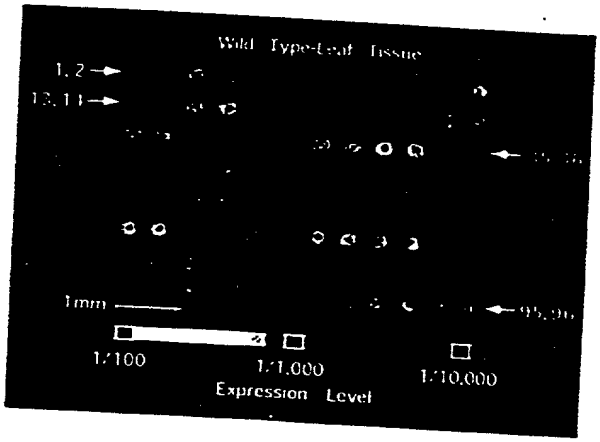
Wild Type-Leaf Tissue
1.2
13.14
1mm
1/100
1/1,000
1/10,000
Expression Level

Fig. 5B

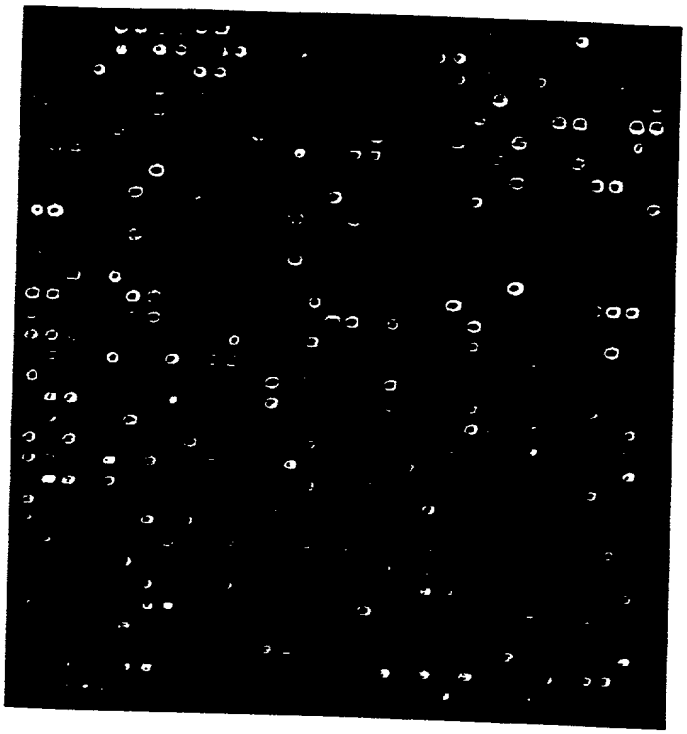

Fig. 6

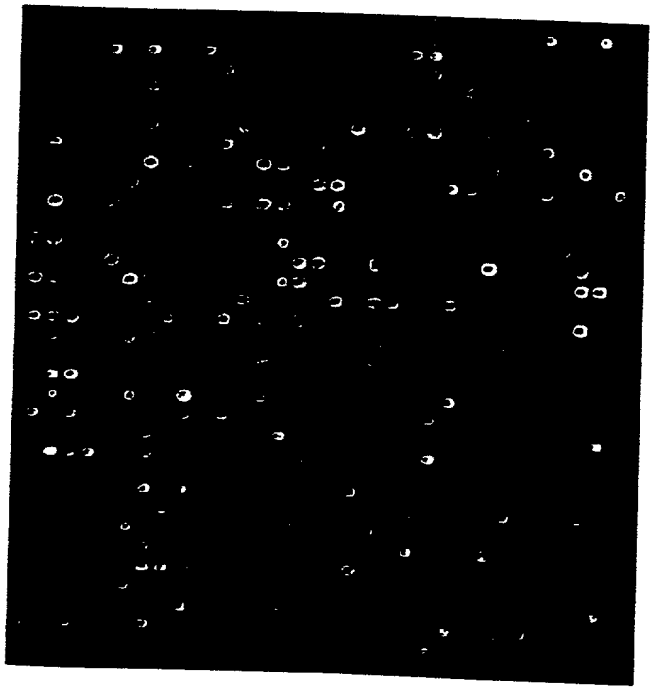

Fig. 7

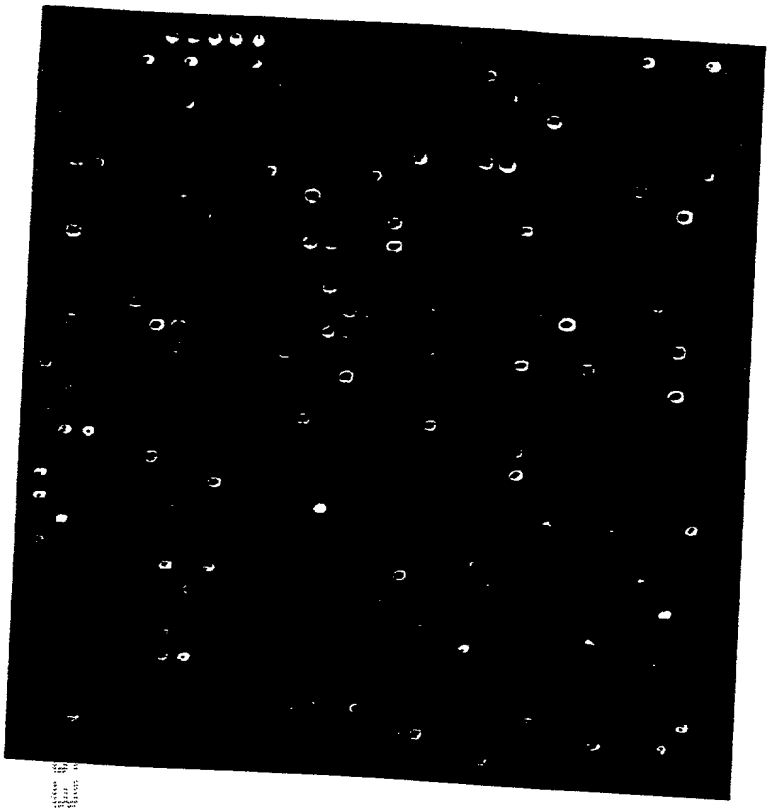

Fig. 8

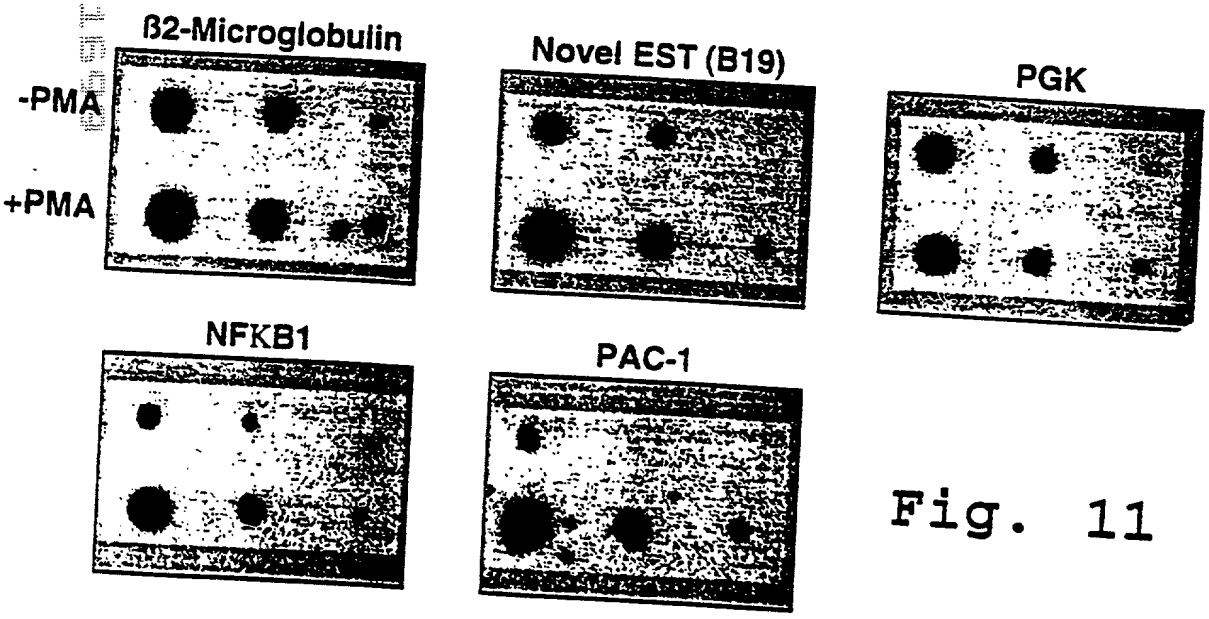

Fig. 11

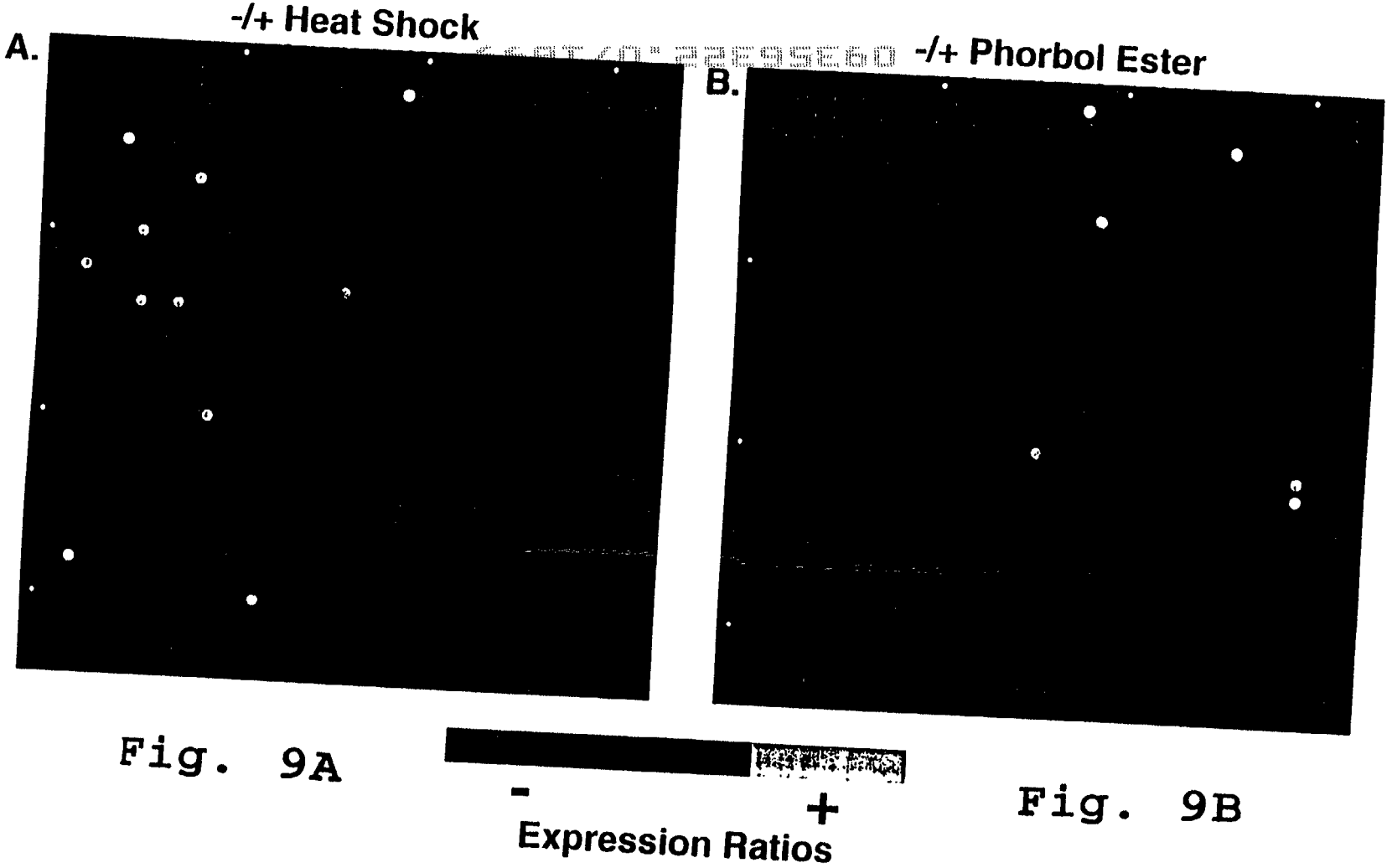

Fig. 9A

Fig. 9B

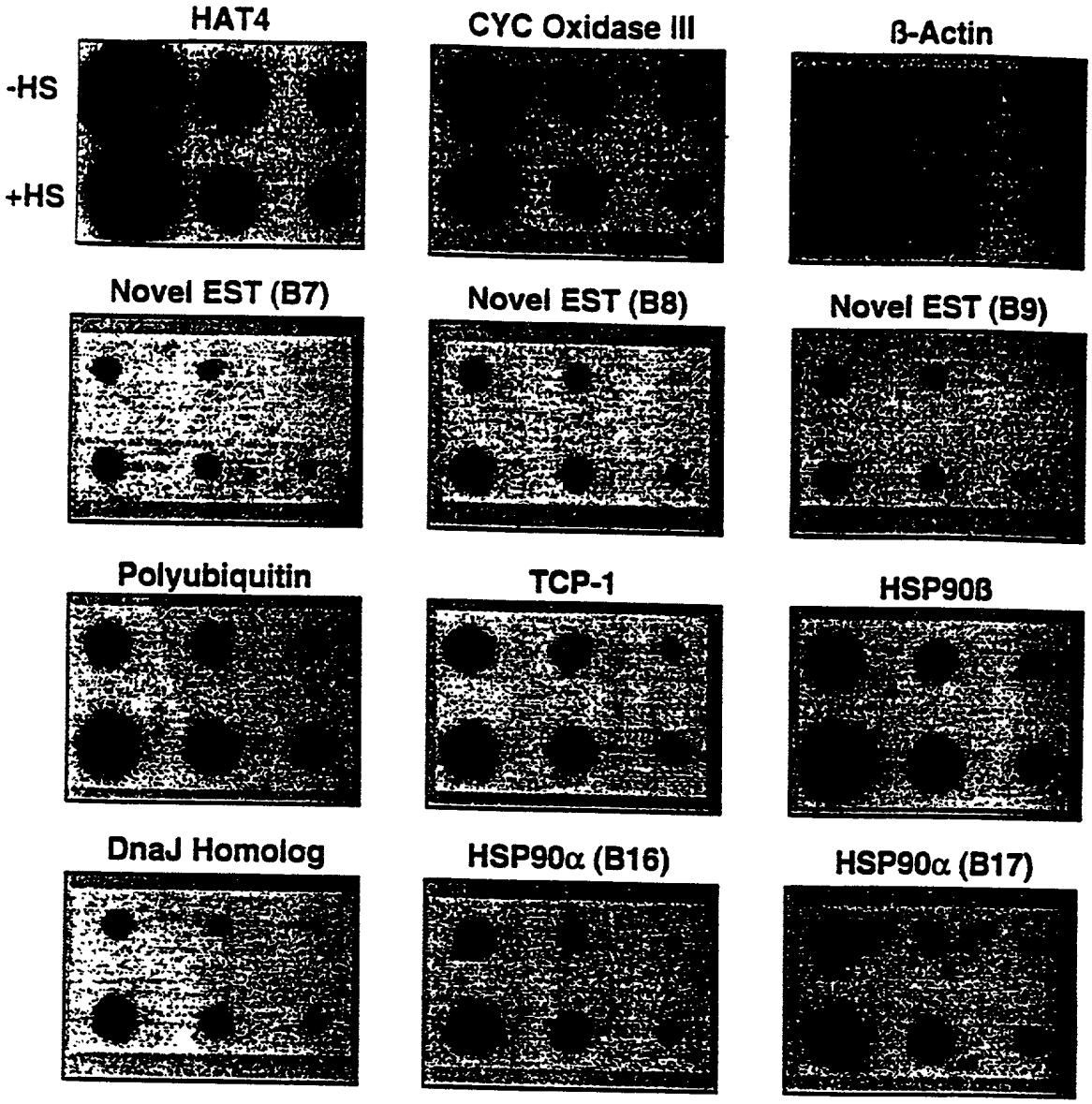
HAT4
CYC Oxidase III
ß-Actin
-HS
+HS
Novel EST (B7)
Novel EST (B8)
Novel EST (B9)
Polyubiquitin
TCP-1
HSP90ß
DnaJ Homolog
HSP90α (B16)
HSP90α (B17)

Fig. 10

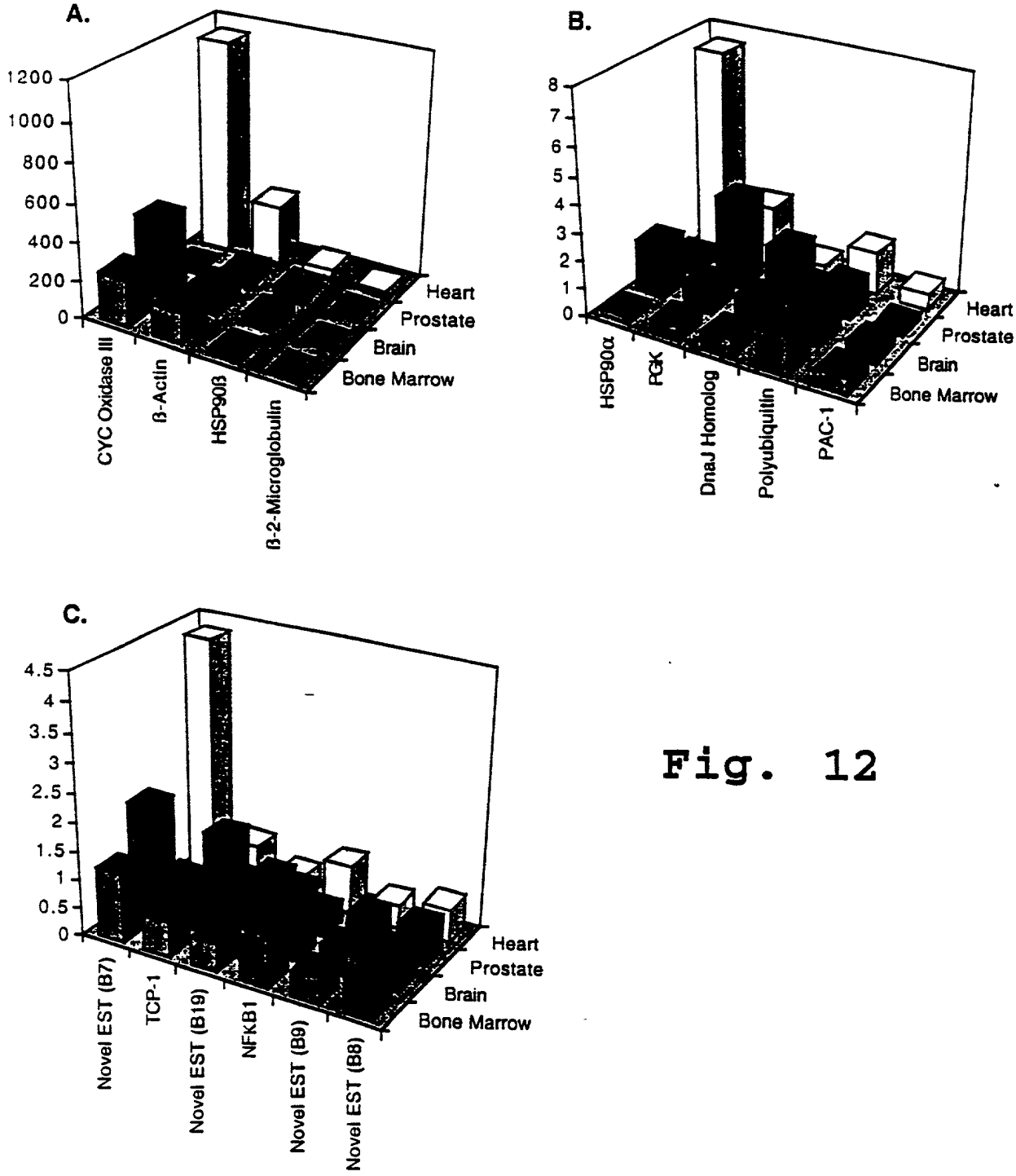
A.
1200
1000
800
600
400
200
0
CYC Oxidase III
β-Actin
HSP90ß
ß-2-Microglobulin
Heart
Prostate
Brain
Bone Marrow
B.
8
7
6
5
4
3
2
1
0
HSP90α
PGK
DnaJ Homolog
Polyubiquitin
PAC-1
Heart
Prostate
Brain
Bone Marrow
C.
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Novel EST (B7)
TCP-1
Novel EST (B19)
NFKB1
Novel EST (B9)
Novel EST (B8)
Heart
Prostate
Brain
Bone Marrow

Fig. 12

Patent Application Declaration and Power of Attorney

I HEREBY DECLARE THAT:

My residence, post office address, and citizenship are as stated next to my name in PART A on page 2 hereof.

I believe I am the original, first, and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed) of the subject matter which is claimed and for which a patent is sought on the invention entitled

Method for Determining the Relative Abundances of Polynucleotide Sequences

the specification of which:

- ☒ is attached hereto.
- ☐ was filed on _____ as Attorney Docket No. _____.
- ☐ was filed on _____ as Application Serial No. _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed in PART B on page 2 hereof and have also identified in PART B on page 2 hereof any foreign application for patent or inventor's certificate having a filing date before that of the application of which priority is claimed.

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed in PART C on page 2 hereof and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application. If the patent application identified above is a continuation-in-part application, I acknowledge the duty to disclose material information as defined in 37 CFR §1.56(a) which occurred between the filing date of the prior application and filing date of the continuation-in-part application which discloses and claims subject mater in addition to that disclosed in the prior application (37 CFR §1.63(d)).

I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint the following as my attorney or agent with full power of substitution to prosecute this application and transact all business in the United States Patent and Trademark Office connected therewith:

Peter J. Dehlinger	Reg. No. 28,006	Judy M. Mohr	Reg. No. 38,563
Carol A. Stratford	Reg. No. 34,444	Susan T. Evans	Reg. No. 38,443
Vincent M. Powers	Reg. No. 36,246	LeeAnn Gorthey	Reg. No. 37,337
Charles K. Sholtz	Reg. No. 38,615		

whose mailing address for this application is: Dehlinger & Associates
P.O. Box 60850
Palo Alto, CA 94306

See Page 2 attached, signed, and made a part hereof.

Patent Application Declaration and Power of Attorney

PART A: Inventor Information and Signature

Full name of SOLE or FIRST inventor **Tidhar Dari Shalon**

Citizenship U.S.A. Post Office Address 364 Fletcher Drive
Atherton, California 94027

Residence (if different) (same)

INVENTOR'S SIGNATURE: _____ DATE: _____

Full name of SECOND joint inventor, if any **Patrick O. Brown**

Citizenship U.S.A. Post Office Address 76 Peter Coutts Circle
Stanford, California 94305

Residence (if different) (same)

SECOND INVENTOR'S SIGNATURE: _____ DATE: _____

* * * * * * * * * *

PART B: Prior Foreign Application(s)

Serial No.	Country	Day/Month/Year Filed	Priority Claimed
			☐ Yes ☐ No
			☐ Yes ☐ No

PART C: Claim for Benefit of Filing Date of Earlier U.S. Application(s)

Serial No.	Filing Date	Status:					
08/514,875	08/14/95	☐	Patented	☒	Pending	☐	Abandoned
08/477,809	06/07/95	☐	Patented	☒	Pending	☐	Abandoned
08/261,388	06/17/94	☐	Patented	☒	Pending	☐	Abandoned

See Page 1 to which this is attached and from which this Page 2 continues.

PATENT

DECLARATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe the below named inventors are the original, first and joint inventors of the subject matter which is claimed and for which a patent is sought on the invention entitled "Method for Determining the Relative Abundances of Polynucleotide Sequences" the specification of which:

___ is attached hereto.
_X_ was filed on July 30, 1996 as Application Serial No 08/688,488.

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims.

I acknowledge the duty to disclose to the Patent and Trademark Office all information known to me to be material to patentability of the subject matter claimed in this application, as "materiality" is defined in Title 37, Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

PRIOR FOREIGN APPLICATION(S)			Priority Claimed
(Number)	(Country)	(Date Filed)	Yes/No
(Number)	(Country)	(Date Filed)	Yes/No

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose all information known to me to be material to patentability of the subject matter claimed in this application, as "materiality" is defined in Title 37, Code of Federal Regulations, § 1.56, which become available between the filing date of the prior application and the national or PCT international filing date of this application:

08/261,388	June 17, 1994	Abandoned
(Application Serial No.)	(Filing Date)	(Status)
08/477,809	June 7, 1995	Pending
(Application Serial No.)	(Filing Date)	(Status)
08/514,875	August 14, 1995	Pending
(Application Serial No.)	(Filing Date)	(Status)

I hereby direct that all correspondence and telephone calls be addressed to Emily A. Evans, Reg. No. 33,493, Arnold, White & Durkee, P.O. Box 4433, Houston, Texas 77210, (415) 614-4500.

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Inventor's Full Name: Tidhar Dari Shalon

Inventor's Signature: [signature]

Date: 16 Sept 96 Country of Citizenship: USA

Residence Address: 2140 A CAMINO A LOS CERROS
(Include number, street name, city and country)

Post Office Address: MENLO PARK CA 94025
(if different from residence address)

Inventor's Full Name: Patrick O. Brown

Inventor's Signature: [signature]

Date: Sept 26, 1996 Country of Citizenship: USA

Residence Address: 76 Peter Coutts Circle, Stanford CA 94305
(Include number, street name, city, state and country)

Post Office Address:
(if different from residence address)

I HEREBY CERTIFY THAT THIS CORRESPONDENCE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE IN AN ENVELOPE WITH SUFFICIENT POSTAGE AS FIRST CLASS MAIL ADDRESSED TO THE COMMISSIONER OF PATENTS AND TRADEMARKS, WASHINGTON, D.C., 20231, ON:

Date: October 10, 1996 By: Emily A. Evans

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	**T. D. Shalon and P.O. Brown**		
Serial No.:	**08/688,488**	Group Art Unit:	**Unknown**
Filed:	**July 30, 1996**	Examiner:	**Unknown**
For:	**Method for Determining the Relative Abundances of Polynucleotide Sequences**	Attorney Docket No.:	**STFD 009**

ELECTION UNDER 37 C.F.R. §§ 3.71 and 3.73
AND POWER OF ATTORNEY

Honorable Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

The undersigned, being Assignee of the entire interest in the above-identified application by virtue of an Assignment enclosed herewith, hereby elects, under 37 C.F.R. § 3.71, to prosecute the application to the exclusion of the inventors.

The Assignee hereby revokes any previous Powers of Attorney and appoints:

Emily A. Evans, Reg. No. 33,493;
Gerald P. Dodson, Reg. No. 32,787; and,
J. Paul Williamson, Reg. No. 29,600,

each an attorney with the law firm of ARNOLD, WHITE & DURKEE, as its attorney so long as they remain with such law firm, with full power of substitution and revocation, to prosecute the application, to make alterations and amendments therein, to transact all business in the Patent and Trademark Office in connection therewith, to receive any Letters Patent, and for one year after issuance of such Letters Patent to file any request for a certificate of correction that may be deemed appropriate.

Pursuant to 37 C.F.R. § 3.73, the undersigned has reviewed the evidentiary documents, specifically the Assignment to **The Board of Trustees of the Leland Stanford Junior University** referenced below, and certifies that to the best of my knowledge and belief, title remains in the name of the Assignee.

Please direct all communications as follows:

Emily A. Evans, Esq.
ARNOLD, WHITE & DURKEE
P. O. Box 4433
Houston. Texas 77210-4433
(415) 614-4500

ASSIGNEE: The Board of Trustees of the
Leland Stanford Junior University

Date: Oct. 9, 1996

By: Luis R. Mejia
Name: ~~Katharine Ku~~ Luis Mejia
Title: Acting Director, Office of Technology Licensing

ASSIGNMENT:

__X__ Enclosed for recording
____ Previously recorded
Date: _____
Reel: _____
Frame: ____